THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Accessing Health and Health-Related Data in Canada

Citation for published version:
Byerring, AK, Brownell, M, El Emam, K, Fortier, I, Henry, D, Knoppers, BM, Laurie, G, Lemmens, T, Morgan, M, Noseworthy, TW, Saunders, S, Wolfson, M & Zelmer, J 2015, *Accessing Health and Health-Related Data in Canada: The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation*. Council of Canadian Academies. <http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/Health-data/HealthDataFullReportEn.pdf>

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 06. May. 2023

Accessing Health and Health-Related Data in Canada

The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation

Council of Canadian Academies
Conseil des académies canadiennes

Science Advice in the Public Interest

ACCESSING HEALTH AND HEALTH-RELATED DATA IN CANADA

The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation

THE COUNCIL OF CANADIAN ACADEMIES
180 Elgin Street, Suite 1401, Ottawa, ON, Canada K2P 2K3

Notice: The project that is the subject of this report was undertaken with the approval of the Board of Governors of the Council of Canadian Academies. Board members are drawn from the Royal Society of Canada (RSC), the Canadian Academy of Engineering (CAE), and the Canadian Academy of Health Sciences (CAHS), as well as from the general public. The members of the expert panel responsible for the report were selected by the Council for their special competencies and with regard for appropriate balance.

This report was prepared for the Government of Canada in response to a request from the Canadian Institutes of Health Research (CIHR). Any opinions, findings, or conclusions expressed in this publication are those of the authors, the Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, and do not necessarily represent the views of their organizations of affiliation or employment.

Library and Archives Canada Cataloguing in Publication

Accessing health and health-related data in Canada : the Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation.

Includes bibliographical references.

Electronic monograph in PDF format.

Summary: "The Canadian Institutes of Health Research (CIHR) has asked the Council of Canadian Academies to assess timely access to health and social data for health research and health system innovation in Canada." – Provided by publisher.

ISBN 978-1-926522-05-0 (pdf)

1. Medical records–Access control–Canada. 2. Medical records–Law and legislation–Canada. 3. Data protection–Law and legislation–Canada. 4. Medical records–Management–Canada. 5. Privacy, Right of–Canada. 6. Health–Research–Canada. I. Council of Canadian Academies, issuing body II. Council of Canadian Academies. Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, author

R864.A24 2015 651.5'04261 C2015-901476-X

This report should be cited as: Council of Canadian Academies, 2015. *Accessing Health and Health-Related Data in Canada*. Ottawa (ON): The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, Council of Canadian Academies.

Disclaimer: The internet data and information referenced in this report were correct, to the best of the Council's knowledge, at the time of publication. Due to the dynamic nature of the internet, resources that are free and publicly available may subsequently require a fee or restrict access, and the location of items may change as menus and webpages are reorganized.

© 2015 Council of Canadian Academies

Printed in Ottawa, Canada

Council of Canadian Academies
Conseil des académies canadiennes

Canada This assessment was made possible with the support of the Government of Canada.

The Council of Canadian Academies

Science Advice in the Public Interest

The Council of Canadian Academies (the Council) is an independent, not-for-profit organization that supports independent, science-based, authoritative expert assessments to inform public policy development in Canada. Led by a 12-member Board of Governors and advised by a 16-member Scientific Advisory Committee, the Council's work encompasses a broad definition of *science*, incorporating the natural, social, and health sciences as well as engineering and the humanities. Council assessments are conducted by independent, multidisciplinary panels of experts from across Canada and abroad. Assessments strive to identify emerging issues, gaps in knowledge, Canadian strengths, and international trends and practices. Upon completion, assessments provide government decision-makers, researchers, and stakeholders with high-quality information required to develop informed and innovative public policy.

All Council assessments undergo a formal report review and are published and made available to the public free of charge in English and French. Assessments can be referred to the Council by foundations, non-governmental organizations, the private sector, or any level of government.

The Council is also supported by its three founding Member Academies:

The Royal Society of Canada (RSC) is the senior national body of distinguished Canadian scholars, artists, and scientists. The primary objective of the RSC is to promote learning and research in the arts and sciences. The RSC consists of nearly 2,000 Fellows — men and women who are selected by their peers for outstanding contributions to the natural and social sciences, the arts, and the humanities. The RSC exists to recognize academic excellence, to advise governments and organizations, and to promote Canadian culture.

The Canadian Academy of Engineering (CAE) is the national institution through which Canada's most distinguished and experienced engineers provide strategic advice on matters of critical importance to Canada. The Academy is an independent, self-governing, and non-profit organization established in 1987. Fellows are nominated and elected by their peers in recognition of their distinguished achievements and career-long service to the engineering profession. Fellows of the Academy, who number approximately 600, are committed to ensuring that Canada's engineering expertise is applied to the benefit of all Canadians.

The Canadian Academy of Health Sciences (CAHS) recognizes individuals of great achievement in the academic health sciences in Canada. Founded in 2004, CAHS has approximately 400 Fellows and appoints new Fellows on an annual basis. The organization is managed by a voluntary Board of Directors and a Board Executive. The main function of CAHS is to provide timely, informed, and unbiased assessments of urgent issues affecting the health of Canadians. The Academy also monitors global health-related events to enhance Canada's state of readiness for the future, and provides a Canadian voice for health sciences internationally. CAHS provides a collective, authoritative, multidisciplinary voice on behalf of the health sciences community.

www.scienceadvice.ca
@scienceadvice

Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation

Andrew K. Bjerring (Chair), Former President and CEO, CANARIE Incorporated (London, ON)

Marni Brownell, Senior Research Scientist, Manitoba Centre for Health Policy; Associate Professor, Department of Community Health Sciences, College of Medicine, Faculty of Health Sciences, University of Manitoba (Winnipeg, MB)

Brent Diverty, Vice President of Programs, Canadian Institute for Health Information (Ottawa, ON)

Khaled El Emam, Associate Professor, Faculty of Medicine, University of Ottawa; Senior Investigator, Children's Hospital of Eastern Ontario Research Institute; Canada Research Chair in Electronic Health Information, University of Ottawa; CEO, Privacy Analytics Inc. (Ottawa, ON)

Isabel Fortier, Researcher, Maelstrom Research Institute, McGill University Health Centre (Montréal, QC)

David Henry, Professor, Dalla Lana Faculty of Public Health, University of Toronto; Senior Scientist and former CEO, Institute for Clinical Evaluative Sciences (Toronto, ON)

Bartha Maria Knoppers, O.C., O.Q., FCAHS, Canada Research Chair in Law and Medicine; Director, Centre of Genomics and Policy, Faculty of Medicine, McGill University (Montréal, QC)

Graeme Laurie, Professor of Medical Jurisprudence, University of Edinburgh; Founding Director, JK Mason Institute for Medicine, Life Sciences and the Law (Edinburgh, United Kingdom)

Trudo Lemmens, Professor and Scholl Chair in Health Law and Policy, Faculty of Law; Professor, Joint Centre for Bioethics and Faculty of Medicine, University of Toronto (Toronto, ON)

Matthew Morgan, Vice President, Patient Experience and Outcomes, Mount Sinai Hospital; Assistant Professor, Division of General Internal Medicine, Department of Medicine, Faculty of Medicine, University of Toronto (Toronto, ON)

Thomas William Noseworthy, C.M., Professor, Health Policy Management, Department of Community Health Sciences and O'Brien Institute for Public Health, University of Calgary (Calgary, AB)

Stephen Saunders, Senior Executive Consultant and Chief Architect, Healthcare, CGI Group (Eganville, ON)

Michael Wolfson, FCAHS, Canada Research Chair, Population Health Modelling/Populomics, Faculty of Medicine, University of Ottawa (Ottawa, ON)

Jennifer Zelmer, Executive Vice President, Canada Health Infoway (Toronto, ON)

Message from the Chair

The objectives of improving the health and well-being of Canadians and of the health system both require ongoing research and innovation. One of the major requirements for addressing these challenges is the availability of high-quality data, including data on individuals and their encounters with service providers in the health system as well as social data on factors that affect health outcomes. At the same time, individuals have a right to privacy; there is a clear obligation that personal health-related data are kept confidential. Striking an appropriate balance between these two imperatives is of fundamental importance. It is also of great concern to numerous organizations and individuals in every jurisdiction in the world, perhaps none more than those who have a responsibility to act as custodians of the data involved.

Ideally, the organizations and individuals who contribute to this collective effort, whether within a single province or territory or at the national level in a federated jurisdiction like Canada, would constitute a coherent and smoothly operating system with well-defined governance principles and efficient operating procedures that, among other things, would support timely access to health and social data for research and system innovation. This tends not to be the case in Canada. Indeed, those who need access to data must navigate a "complex environment of heterogeneous entities," often including numerous data custodians, privacy offices, and research ethics boards, whose collective governance and operational practices fall short of constituting a well-defined and coherent system.

To address the challenge of providing timely access to health and social data within this context, the Expert Panel was asked, among other things, to identify where the provision of such access could be seen as constituting a "best practice." One particularly noteworthy finding of this report is that many of the "best practice entities" identified here were themselves created as a result of a review of the collective behaviour of the complex environment existing in their particular jurisdiction. In other words, the undertaking of a review by a provincial, territorial, or federal jurisdiction of how well its complex environment addresses *collective* governance responsibilities itself constitutes a best practice.

On behalf of the Panel I would like to thank those who met with us early in the process to help us tackle our charge. I would also like to acknowledge the significant contribution of Council staff to the Panel's work, which would have been impossible without their professionalism, patience, and insight into how we might best make a contribution. Finally, I would like to personally thank the Panel members for their dedication and hard work. I cannot imagine any group of individuals better positioned to help the cause of providing timely access to data for health research and system innovation in Canada. Their report deserves careful consideration.

Andrew K. Bjerring,
Chair, Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation

Acknowledgments

Over the course of its deliberations, the Panel sought assistance from many individuals and organizations that provided valuable evidence, information, and assistance in the development of the report. The report also benefitted from numerous interactions with participants at the International Health Data Linkage Conference held in April 2014 in Vancouver, Canada. Special thanks go to the following: Carolyn Adams, Macquarie University; Judy Allen, University of Western Australia; Jane Badets and Lynn Barr-Telford, Statistics Canada; Daniel Bedard, BORN Ontario; Christiane Bétie, Commission d'accès à l'information du Québec; Pam Bjornson, National Research Council Canada; Charles Burchill, University of Manitoba; Ann Cavoukian, Former Information and Privacy Commissioner, Ontario; Geoff Davis, Department of Health Western Australia; Felicity Flack, Population Health Research Network; David Ford, Swansea University; Karey Iron, Institute for Clinical Evaluative Sciences; Patricia Kosseim, Office of the Privacy Commissioner of Canada; Nancy Meagher, University of British Columbia; Peter Morrison, Statistics Canada; Stephen Pavis, Farr Institute @ Scotland; Daryl Pullman, Memorial University; Parminder Raina, McMaster University; Diana Rosman, Department of Health Western Australia; Mark Smith, University of Manitoba; Merran Smith, Population Health Research Network; and Alan Winter, Genome British Columbia.

The Council wishes to thank the Honourable Elizabeth Dowdeswell, O.C., O.Ont., Lieutenant Governor of Ontario, and Former President of the Council of Canadian Academies, for her guidance and support during the early stages of this assessment.

Project Staff of the Council of Canadian Academies

Assessment Team:	Christina Stachulak, Senior Program Director Aled ab Iorwerth, Research Associate Jennifer Bassett, Research Associate Weronika Zych, Program Coordinator
With assistance from:	Haryanto Darmawan, Consultant Thomas Archibald, Consultant Edward Dove, Consultant Carolyn Brown, Editor Clare Walker, Editor and Copyeditor Accurate Design & Communication, Report Design

Report Review

This report was reviewed in draft form by the individuals listed below — a group of reviewers selected by the Council of Canadian Academies for their diverse perspectives, areas of expertise, and broad representation of academic, industrial, policy, and non-governmental organizations.

The reviewers assessed the objectivity and quality of the report. Their submissions — which will remain confidential — were considered in full by the Panel, and many of their suggestions were incorporated into the report. They were not asked to endorse the conclusions, nor did they see the final draft of the report before its release. Responsibility for the final content of this report rests entirely with the authoring Panel and the Council.

The Council wishes to thank the following individuals for their review of this report:

Mark Dermer, Family Physician and Managing Associate, Central Ottawa Family Medicine Associates (Ottawa, ON)

David V. Ford, Professor of Health Informatics, College of Medicine, Swansea University (Swansea, Wales)

Patricia Kosseim, Senior General Counsel and Director General, Legal Services, Policy, Research and Technology Analysis Branch, Office of the Privacy Commissioner (Ottawa, ON)

Adrian Levy, Professor and Head, Department of Community Health and Epidemiology, Dalhousie University (Halifax, NS)

David Loukidelis, Q.C., Former Information and Privacy Commissioner for British Columbia; Former Deputy General of British Columbia; Former Adjunct Professor, Freedom of Information and Privacy Law, Faculty of Law, University of Victoria; Former Adjunct Professor, Freedom of Information and Privacy Law, Faculty of Law, Thompson Rivers University (Edmonton, AB)

Stan Matwin, Canada Research Chair and Professor, Director, Institute for Big Data Analytics, Faculty of Computer Science, Dalhousie University (Halifax, NS)

Ted McDonald, Professor, Department of Economics, University of New Brunswick (Fredericton, NB)

Nancy Meagher, Executive Director, Population Data BC (Vancouver, BC)

Noralou Roos, C.M., FRSC, FCAHS, Professor, Faculty of Health Sciences, University of Manitoba (Winnipeg, MB)

Charlie Schick, Director, Business Development, Healthcare, Atigeo LLC (Bellevue, WA)

Fiona Stanley, Patron, Telethon Kids Institute; Distinguished Research Professor, The University of Western Australia; Vice Chancellor's Fellow, The University of Melbourne (Perth, Australia)

Don Willison, Associate Professor, Institute for Health Policy, Management and Evaluation, University of Toronto (Toronto, ON)

Eric R. Wright, Professor, Department of Sociology, College of Arts and Science and School of Public Health, Georgia State University (Atlanta, GA)

Glenda Yeates, Former Deputy Minister, Health Canada (Ottawa, ON)

The report review procedure was monitored on behalf of the Council's Board of Governors and Scientific Advisory Committee by **Lorne Babiuk, O.C., FRSC, FCAHS,** Vice President (Research), University of Alberta (Edmonton, AB). The role of the report review monitor is to ensure that the Panel gives full and fair consideration to the submissions of the report reviewers. The Board of the Council authorizes public release of an expert panel report only after the report review monitor confirms that the Council's report review requirements have been satisfied. The Council thanks Dr. Babiuk for his diligent contribution as report review monitor.

Janet Bax,
Interim President Council of Canadian Academies

Executive Summary

INTRODUCTION

Canadians care deeply about health care — for themselves, their families, and their communities. Ensuring that the health-care system can deliver the best possible care depends fundamentally on research into system innovation and health and social well-being. This research depends on the availability of high-quality data.

The volume and variety of data relevant to such research have increased exponentially in recent years. Each patient interaction with a physician, a pharmacist, a laboratory technician, or hospital staff generates data. Social and environmental data are highly relevant to health research because they are vital for providing a complete picture about factors that affect the lives and health of Canadians. The research community, including health system innovators in hospital and government offices as well as academic researchers and clinicians, views these data as a critical resource. It recognizes the enormous potential of using health and health-related data in privacy-sensitive ways to reveal factors that can affect health and well-being, and discover interventions that can improve health outcomes.

Despite these benefits, working with the data on which the research is based can be challenging. Some challenges are technical, such as the use of different standards in different jurisdictions to record important data. Others are related to privacy concerns: access to health data for research carries the risk that personal data could be released, whether inadvertently or intentionally. The greatest challenges, which are indeed barriers to beneficial research, are institutional. These include the application of differing, and in some instances overly cautious, interpretations of privacy legislation, and complex and lengthy approval processes that impede researchers' access to data.

The primary, overarching challenge in Canada, as in other jurisdictions, is to meet two fundamental goals at the same time: to enable access to health and health-related data for research that is in the public interest, on the one hand, and to respect Canadians' privacy and maintain confidentiality of their information when it is used for research, on the other. Innovative organizations and less formal collaborative undertakings are finding ways to meet these goals. They are instituting governance models and practices that further scientifically sound, ethically robust research and respect privacy, while using technology in innovative ways to provide data access in a timely and confidentiality-preserving manner.

Charge to the Panel

In 2013, the Canadian Institutes of Health Research (CIHR; the Sponsor) asked the Council of Canadian Academies (the Council) to respond to the following charge:

> *What is the current state of knowledge surrounding timely access to health and social data for health research and health system innovation in Canada?*

The charge also included five sub-questions:

- *What is known about how to address technological and methodological challenges (such as variable data quality and comparability) associated with linkage of health and social data from various sources and across jurisdictions?*
- *What is known about the benefits, risks and barriers to timely access to health and social data for health research and health system innovation in Canada?*
- *What are the ethical, legal, and social implications of timely access to such data?*
- *What are best practices for improving access to such data for researchers while ensuring appropriate privacy safeguards and also taking full advantage of the digital data revolution?*
- *What are best practices in Canada and internationally for governance frameworks that facilitate access to such data and maintain public trust in the research enterprise?*

To address these questions, the Council formed the Expert Panel on the Timely Access to Health and Social Data for Health Research and Health System Innovation (the Panel), which comprised 14 Canadian and international experts from the health-care sector, academia, and industry. Panel members had experience as data custodians, researchers, managers of health research organizations, or in legal aspects of health research.

At the outset of the assessment, the Sponsor gave further direction on interpreting and refining the charge. First, the Sponsor defined *timely access* as access granted within four months of submitting a data request to an organization responsible for providing the data. Second, the assessment should concern only public interest research (i.e., research conducted by public bodies and/or supported by public funds). Thus, health and health-related data used by private, commercial

companies were excluded. Third, the assessment should identify best practices in Canada and other countries for timely access to data that can be linked and integrated for research purposes, rather than exploring barriers to accessing data in general. Finally, the Panel's work should encompass all types of health data related to publicly funded research, ranging from administrative health data to genomic data. The Panel understood the inclusion of social data as data on non-medical determinants of health such as health behaviours, living and working conditions, personal resources, and environmental factors, and hence uses the term *health-related data*.

Methodology for Identifying Best Practices

The identification of best practices was the first issue addressed because of broad implications for the Panel's overall approach. The Panel looked for organizations, institutions, programs, or other entities that had been especially successful in meeting the twin goals of enabling timely access and protecting privacy. It selected six entities, three from Canada and three from other jurisdictions with similar legal and social systems:

- Manitoba Centre for Health Policy (MCHP)
- Ontario – Institute for Clinical Evaluative Sciences (ICES)
- Ontario – Better Outcomes Registry and Network (BORN)
- Wales Secure Anonymised Information Linkage Databank (SAIL)
- Data Linkage Western Australia (Data Linkage WA)
- Farr Institute @ Scotland[1]

These "best practice entities" are mandated, in some cases under legislation, to receive data from encounters in the health-care system and to provide access for public interest research. They all succeed in providing access within a four-month timeframe and share four common principles:

- **Enabling appropriate use of data** to enhance public well-being;
- **Managing risk** by identifying the range of risks involved in providing data access and minimizing those risks where possible, while acknowledging that risks cannot be entirely eliminated;
- **Respecting privacy** to reassure citizens that risks to their core personal interests are kept to an absolute minimum; and
- **Maintaining public trust** by providing evidence of trustworthiness, including using data appropriately and demonstrating the social value of the resulting research.

1 The Farr Institute @ Scotland builds on the success of the Scottish Informatics Programme (SHIP), which ran from 2009 to 2013.

The practices highlighted by the Panel reflect both a literature review as well as the practices of the six entities. The Panel found many examples of *good practice*, including approaches for dealing with legal and ethical considerations. These are highlighted in Finding 4 below. However, in accordance with the charge, the Panel only identified *best practice* related to governance that could be put in place to enable access to health data. Best practices for governance are discussed in Finding 5. *Good practice* and *best practice* are defined in the glossary that accompanies the full report.

The Panel also examined many other organizations that provide access to health and health-related data, or that play a special role in analyzing such data, including, among others, Statistics Canada, Statistics Netherlands, the U.S. National Institutes of Health, Population Data BC, the Canadian Network for Observational Drug Effect Studies (CNODES), and the Canadian Institute for Health Information (CIHI). Many insights can be drawn from these innovative organizations about striking an appropriate balance between respecting privacy and providing timely access to data.

In addition to identifying best practice organizations, the Panel reviewed evidence on how other organizations in Canada and around the world enable access to health and health-related data for research. It drew from published literature, conference proceedings, and online reports, including a key OECD report (*Strengthening Health Information Infrastructure for Health Care Quality Governance*) that summarizes the use of health data in a range of countries.

KEY FINDINGS

The Panel's findings fall into five categories that roughly correspond to the five sub-questions of the charge: technological and methodological challenges, benefits, risks, legal and ethical considerations, and governance.

Overall, the Panel found data-intensive research has both clear benefits and risks. Striking the right balance can be achieved through good governance that demonstrates respect for legal and ethical considerations, and for the people whose data are being used.

1. Technological and Methodological Challenges of Access to Health Data

For effective research with health and health-related data, disparate sources of data must be brought together. Providing these data in an "analysis-ready" format, thereby allowing statistical relationships or patterns to be derived, is a central methodological challenge.

The full potential of Canada's health and health-related data can only be realized if the data are made ready for analysis. However, much of the data with the greatest potential for research are collected for other purposes, such as administration of health care services. To be used for research, these data need to be transformed into specific forms and formats — predominantly statistical ones. As electronic health records (EHRs) become increasingly prevalent, it will be more efficient to anticipate and design into these data the capacity to support secondary use rather than to retrofit after computer systems for EHR recording have already been designed.

EHRs and health-care encounter data inherently involve many disparate sources of data, from hospitalizations to lab tests. Thus, for research as well as effective patient care, it is necessary to bring different data sets together. The key difference is that for patient care, the focus is on a single patient, while for research, the focus is on large samples of individuals, where any given individual's identity is irrelevant. As a result, research-oriented data sets may be from the same province/territory, multiple provinces and territories, or multiple countries.

To be compared or combined and used meaningfully in statistical analysis, data elements must be *harmonized*. The best approach for harmonization involves the development of standard terminologies, questionnaires, measurements, and protocols (i.e., *prospective harmonization*). But this approach may be too challenging, time-consuming, or labour-intensive; or an underlying consensus on how to define or measure a given variable may be absent. In these situations, *retrospective harmonization* can be attempted. Tools are available to help determine whether similar inferences can be drawn from variables across different studies.

Data linkage allows different types of information for one individual to be brought together. It can be challenging if (i) unique identifiers are not available for all individuals in a data set, or (ii) data have been strongly de-identified.[2] To overcome the first challenge, probabilistic methods can be used to link records. The simplest solution to the second challenge is to link the data prior to de-identification, if possible. Databases do not always need to be linked permanently. The link can be destroyed after the research is completed, and/or kept completely separate by implementing the *separation principle*.

2 De-identification is the act of minimally perturbing individual-level data to decrease the probability of discovering an individual's identity. It involves masking direct identifiers (e.g., name, phone number, address) as well as transforming indirect identifiers that could be used alone or in combination to re-identify an individual (e.g., birth dates, geographic details, dates of key events).

Pooling of similar data from several populations is often used to increase the sample size for a study. *Bona fide* pooled data analysis involves physical transfer of individual-level data to a central server, where the data are then analyzed as they would be if they were from the same study (with statistical adjustments if needed). In many important cases in Canada, restrictive interpretations of privacy and other laws have hindered pooling of individual-level data from different provinces. Therefore, approaches that avoid the need to pool individual-level data have been developed. One of these approaches (used by CNODES) involves statistical analyses of harmonized, individual-level data at each study site, followed by pooling of the (non-confidential) summary statistics to obtain an overall result. Another, provided by DataShield, uses sophisticated iterative techniques to mimic a pooled analysis of data from individual participants, when, in reality, the data always remain with their original data custodian.

2. Benefits of Access to Health Data

Evidence shows that timely access to data enables significant high-quality research that can have far-reaching effects for health care and the overall health of Canadians.

Timely access to health and health-related data enables significant high-quality research, which identifies risk factors for various health and social outcomes, and determines health interventions with the most beneficial effects. The knowledge gained from this research is fundamental for improving health generally, and maintaining high quality health care. Recent Canadian studies with significant clinical or public health implications have demonstrated the benefits of research using health and health-related data. For example, analysis of data from the Canadian Community Health surveys by researchers at ICES and Public Health Ontario led to the development of a Life Expectancy Calculator that helps Ontarians understand the effect of certain behaviours on their life expectancy. Researchers at MCHP used record linkage to show that low socio-economic status affects educational achievement much more than previously thought. CNODES analysis of hospital data from across Canada showed that seniors over age 65 were five times more likely than the rest of the population to be hospitalized for adverse drug reactions due to specific risk factors such as drug interactions.

3. Risks of Access to Health Data

The risk of potential harm resulting from access to data is tangible but low. The level of risk can be further lowered through effective governance mechanisms.

While there are clear benefits of research using individual Canadians' personal health and health-related data, there are also risks. These can include accidental release of identifiable data, to the public or unauthorized researchers, when proper security and privacy protocols are not followed (e.g., through loss of computer equipment); illicit access to identifiable data (e.g., through hacking); and inadvertent access to identifiable data by those working inside data organizations.

While these types of breaches have occurred during research projects, breaches rarely happen at institutions with databases set up specifically for maintaining large volumes of health and health-related data for research and administrative purposes. They are much more likely to occur when researchers or employees are accessing data directly from health-care centres. Importantly, there are no examples of breaches at the six best practice entities identified by the Panel.

In many cases, the data that researchers access from secure facilities are de-identified. However, re-identification remains a concern. The Panel found that best practices in de-identification can lower the risk of re-identification to acceptable levels. Although health data breaches can cause serious harm, the risk of a breach actually occurring in the context of research is low, particularly if effective governance mechanisms and protocols are in place and respected by care providers, researchers, and data custodians.

4. Legal and Ethical Considerations of Access to Data

Timely access to data is hindered by variable legal structures and differing interpretations of the terms *identifiable* and *de-identified* across jurisdictions. Instead of rigidly classifying data as either identifiable or non-identifiable, it is useful to view de-identification as a continuum and to adjust access controls accordingly.

In enabling access to data for research, the benefits of research, as well as the range of risks, need to be weighed. Canadian research projects demonstrate that beneficial research can be advanced while maintaining confidentiality of sensitive personal information. Yet, access to data and successful data-based research is not uniform across Canada because of (i) the lack of consistency and clarity in Canada's ethical and legal framework, and (ii) differing interpretations of key terms and issues across the country.

While federal and provincial/territorial laws generally allow researchers to access data that do not include "identifiable information," this term is not always defined precisely. This makes it confusing to base data sharing guidelines on the notion that "non-identifiable data" can be used freely. As well, data custodians may interpret their legal duty to protect privacy as precluding access. Laws on sharing data across provinces/territories and countries differ or are lacking, which can also make researchers and research ethics boards (REBs) uncertain as to whether data can be shared.

This lack of legal clarity has contributed to cautious and conservative interpretations of allowable access in many Canadian organizations. While the law provides specific limits for data custodians, it is less specific in other areas. And although provincial and federal laws lay out broad rules about when and how data can be used or shared, often they are silent on specific questions about *whether* data should be so used in specific settings. This means that data custodians often face an asymmetry — there are clear sanctions if there is a data breach when they are in charge, but no benefit to them if their release of data for *bona fide* research generates important public benefits. This asymmetry supports a tendency to *not* grant access, even if access would be acceptable within their legal frameworks.

A number of good practices for addressing legal and ethical issues are summarized below.

Good Practices – Legal and Ethical Considerations

Appropriate access controls for differing levels of de-identification: Because data may be fully identifiable (i.e., no identifiers removed), mildly de-identified, or strongly de-identified, the Panel did not single out one specific process for dealing with de-identified data. Rather, a good practice is to use the degree of de-identification to determine the circumstances under which the data may be made accessible for research purposes (i.e., increase access control as identifiability increases).

Rules governing sharing of identifiable data for research purposes: Maintaining a set of rules that govern the sharing and use of fully identifiable or partially de-identified data for research purposes is a good practice. Examples of such rules are as follows:

- **Data are held at designated research entities:** In some provinces, the legislation designates specific entities that may receive health and health-related data without consent for research purposes, acknowledging that establishing such centres is in the public interest.
- **Research meets approval criteria:** To ensure privacy is respected, and to clearly delineate the requirements for access to identifiable data without consent, good practice suggests showing that the research serves the public interest; obtaining consent is impracticable; identifiable data are necessary to the research project; and physical, electronic, software, and all other security measures are appropriately calibrated to protect the data and to sanction any misuses.
- **Researchers sign researcher-custodian agreements:** To ensure that researchers are accountable for protecting data confidentiality, good practice suggests that full and explicit data transfer agreements between researchers and custodians are needed for each research project.

Risk management strategies: The Tri-Council Policy Statement, which governs ethical research in Canada, recognizes that risk cannot be eliminated but should be considered proportionately. Good practice suggests incorporating risk management in all aspects of governance, including ethical governance.

Establishment of dedicated governance: Whatever the applicable law in a given jurisdiction, it may be open to a considerable range of interpretation. A dedicated governing body is a good practice that could, for example, establish reasonable processes to de-identify data, as well as ensure respect for overall legal and ethical principles.

5. Governance

Evidence demonstrates that a shift is occurring among leading entities from a "data custodianship" model to a "data stewardship" model. Central to the success of this shift is the adoption of good governance practices, specifically in privacy governance, research governance, information governance, and network governance.

The Panel found a marked shift among the six best practice entities from a "data custodianship" model, in which holding and securing data are emphasized to the exclusion of other considerations, to a "data stewardship" model, in which enabling access is a core institutional objective proportionately balanced with protecting privacy. The balance is achieved through good governance, which encompasses the definition of an entity's purpose, objectives, values, and policies.

Addressing the question of providing timely access to health data for research is particularly challenging in Canada as the many institutions, organizations, programs, and activities that deal with health and health-related data are only loosely coordinated. They are best thought of as a "complex environment of heterogeneous entities," the parts of which were not designed to work in concert with one another as a system with a common overall purpose.

Over time, coordination, consistency, and overall effectiveness of the "complex environment" could be achieved through the adaptation of the pre-existing entities. Alternatively, the responsible governments could carry out a broad review and subsequent redesign of their system, comparable to that undertaken in Wales or Scotland.

Along with the other organizations in each of their jurisdictions, the six best practice entities share a *collective* responsibility for addressing several cross-cutting aspects of governance. The Panel found four particularly relevant aspects of governance: privacy governance, information governance, research governance, and network governance. When considered together, these four aspects provide a reasonable framework for examining how the complex environment as a whole governs access to health and health-related data for research.

Privacy Governance

Privacy governance involves monitoring the specific risk to privacy posed by data access by researchers and protecting data confidentiality. Such governance may involve specialized knowledge of technology, privacy law, ethics, and statistical methods.

This aspect of governance ensures appropriate use of confidential data in carefully defined circumstances and under specific conditions. Principles may be put in place to guide access to and protection of personal confidential data.

The six best practice entities have dedicated processes to evaluate privacy concerns when enabling data access. For example, MCHP operates within the context of Manitoba legislation where the Health Information Privacy Committee is responsible for approving health research projects that use personal health information held by a government department or agency. In Wales, SAIL's Information Governance Review Panel (IGRP) is dedicated to privacy review, which ensures appropriate de-identification of data and addresses research ethics concerns. In Ontario, data from the BORN database are certified to indicate that they are de-identified in an approved way, and data from ICES are governed by internal procedures set in consultation with the Information and Privacy Commissioner of Ontario.

Best Practices – Privacy Governance

Dedicated Privacy Evaluation: The best practice entities have developed dedicated processes (parallel to REBs) that specifically evaluate privacy concerns when enabling data access.

Research Governance

The processes and entities that govern the research enterprise in Canada face special challenges in connection with research using health and health-related data. While research governance entails many aspects, the panel chose to focus on the REB process. Of particular importance are the requirements for research projects to be approved in advance by an REB, and for data access requests to be approved, often through a separate process. Timeframes for these approvals vary widely across organizations and jurisdictions in Canada, ranging from months to years. Ethics approval for research projects that involve more than one centre or more than one province/territory, in particular, can involve time-consuming (and duplicative) approval processes.

This issue has been addressed in New Zealand and Wales, as well as in two Canadian provinces, through a reduced number of REBs. Alberta decreased its REBs from six to three. Newfoundland and Labrador has created a central research ethics authority that oversees ethics review but can also approve reviews from other boards within and outside the province, thereby avoiding duplicate reviews.

Another challenge arises when REBs and other boards are inconsistent in interpreting ethical and legal guidelines, for example, regarding what constitutes identifiable information. To overcome this potential problem, many countries and Canadian provinces have established a separate review process for data access requests (e.g., HIPC in Manitoba).

Best Practices – Research Governance

Harmonized REB process: To minimize the number of approvals when performing cross-subject or cross-jurisdictional research — and therefore to improve timeliness — certain jurisdictions such as Alberta, New Zealand, and Wales have harmonized the REB process.

Information Governance

Information governance addresses how information is handled within an organization or among organizations. It covers data organizations and their employees, researchers accessing data, and public input. This aspect of governance is concerned with enabling access to data, and doing so within a reasonable timeframe. The best practice entities have made enabling access one of their central purposes, and, as a result, are moving towards a culture of data stewardship.

Physical and technical measures are also required to enable access to data. However, approaches to data access are on a spectrum, with progressively greater security and precautions as data are less aggressively de-identified. Some organizations allow researchers to access data sets containing identifiable information only at secure locations, often called "safe havens" (e.g., MCHP, ICES, Statistics Canada), or through secure internet connections (e.g., Statistics Netherlands, Population Health Research Network in Australia). For both identifiable and de-identified data, however, the researcher is typically bound by confidentiality agreements and/or the research is subject to pre-approval. Data that are very strongly de-identified may be made publically available by large entities. For example, Statistics Canada provides public-use files for data that are rendered non-identifiable within the meaning of the *Statistics Act*. In some cases, however, these highly de-identified data are much less valuable for research.

Linking data sets across organizations could raise the possibility that many employees can access large amounts of identifiable data. To address this, institutional structures can be established to minimize the risks. One way to manage employee access, referred to as the *separation principle*, is to separate data into a demographic component (with identifying information such as name, address, etc.) and a content component (with information such as medicines prescribed, test results, etc.). This prevents any given individual from seeing both components. The separation principle can be observed by using an external organization to deal with identifying information or by managing all data internally but ensuring that identifying data and content data are administratively — and sometimes physically — separated.

A critical element of any information governance model is the determination of an "acceptable" level of risk, which relies on the development of a method to characterize risk. To address this need in the context of product safety, the European Commission has developed a risk assessment matrix. The Farr Institute @ Scotland has adopted a "proportionate governance" approach in which the level of scrutiny for a data linkage request depends on the level of risk that it entails.

To analyze data across provincial or national boundaries, innovative methods are being undertaken. Through CNODES, data on drug effects are analyzed in each province using standardized methods, and a meta-analysis is conducted on a national level to determine the scale of effects for Canada. Other suggestions include encryption of raw data, and security of core identifiable data with release of summary statistics for analysis via the internet. Various techniques are used to ensure that the system works effectively and efficiently. Common features include (i) adoption of privacy management programs, (ii) adoption of an effective risk management framework, and (iii) adoption and documentation of a "reasonable" process of de-identification.

In summary, application of information governance practices can effectively deal with the public concern of risks such as inadvertent access to data and accidental release of data through loss or theft.

Best Practices – Information Governance

Data access: Certain entities successfully maintain data confidentiality through safe havens and/or encrypted access. Key features of a well-functioning safe haven include mechanisms to approve researchers, robust internal and external monitoring and oversight, and ongoing review of governance arrangements over time.

Enabling data use: Appropriate provision of data to researchers is central to the best practice entities. For example, the mission statements of the Farr Institute @ Scotland, SAIL, Data Linkage WA, ICES, and MCHP clearly lay out that enabling appropriate use of data is a core purpose of their organization.

Privacy management: Entities have developed comprehensive researcher-custodian agreements to ensure that researchers maintain the confidentiality of the information that they receive.

Appropriate institutional structure, respecting separation principle: Entities that use the separation principle have minimized the risk of inadvertent and inappropriate access to data by staff.

De-identification of data: Robust de-identification techniques that have met legal standards (i.e., de-identification is "reasonable") have made it possible to reduce the risk of re-identification to a level that is appropriate for a given access mode (and its accompanying security controls). These include practices to ensure that de-identification is documented, transparent, and meets statistical thresholds for re-identification risk while maintaining data utility.

Technology's role in enabling access to and safeguarding data: New technologies can be adopted and developed to improve the safeguards on confidentiality. Given the central importance of technology, it is critical to have individuals with knowledge of its importance involved in governance.

Acceptable level of risk: The European Commission has developed a systematic method for characterizing risk. Scotland has integrated a proportionate approach to risk in its governance system.

Network Governance

The creation of collaborative research networks, potentially involving not just a circle of researchers but also other stakeholders such as data custodians and funding agencies, has the potential to maximize social benefits flowing from data-oriented research. Given Canada's complex and heterogeneous set of actors and stakeholders, governance to create and maintain these networks is vital for standardizing data collection and developing policies for national and international data sharing.

Among the benefits of building a research network is that it may be the only way to amass enough data to conduct a study. A by-product of network-driven collaborations is that definitions and standards must be defined in advance to make the data involved comparable. Thus, networks are a central contributor to standardization and harmonization. Standardization has been a core function of CIHI and the WHO, whose boards and committees represent another type of network, composed of individuals who may have different research interests and diverse professional backgrounds but who share the common goal of developing national or international standards. Standardization is also a main objective of Statistics Canada.

It is also important that networks develop standardized data security protocols. Genetics initiatives such as the International Cancer Genome Consortium (ICGC) are among the most advanced in their successful development of policies for international sharing of individual-level data.

Networks may play a role in mitigating inaccuracy of research results. Analysis of large data sets involves complex statistics, and results can be erroneous if there is a lack of expertise in this area. Networks can put in place standards for statistical analysis and share information about issues in statistical methods. If incorrect research conclusions are publicly released, networks can act to address these both within the scientific community and vis-à-vis the public.

Best Practices – Network Governance

Data harmonization: To enable prospective data harmonization, entities such as the WHO and CIHI have put standards in place prior to data collection.

Distributed analysis: When it is not possible to pool individual-level data, other models, such as CNODES in Canada and DataShield in Europe, have been successful in enabling statistical analysis across jurisdictions.

Multinational sharing: When legal systems differ, methods have been developed to further research by multinational consortia such as the International Cancer Genome Consortium.

A "Best Practice" Governance Model

The jurisdictions involving some of the best practice entities chosen by the Panel, in particular Wales and Scotland, consciously decided to redesign their entire complex environment of entities involved in health and health-related data for research. Their aim was to prevent overlap, duplication, and confusion, and more effectively address the challenges of privacy, information, research, and network governance. For example, one element of Scotland's good governance framework is a mechanism based on proportionate governance to ensure that data access requests with lower risks receive lighter touch governance. Another element is an "account of responsibilities" of key actors and decision-makers. In contrast, the new system in Wales incorporates all governance into a single governance review panel. Clarifying the responsibilities of key entities in Canada's complex environment could be a positive step in enabling timely access to health and health-related data for research. Chapter 5 of the report summarizes the roles of different groups (e.g., researchers, data custodians, policy makers) and governing bodies (e.g., REBs, privacy monitoring boards) in overseeing various aspects of governance and provides examples of entities that are following best practice by successfully performing these roles.

CONCLUSION

To ensure that Canadians continue to have access to high-quality health care, and benefit from effective health policies, the country's health researchers and system innovators need to make effective use of health and health-related data, including administrative health and social data. This need will increase in the future as technology continues to develop and digitized data such as EHRs become ever more abundant.

However, timely access to health and health-related data for research varies across Canada. While some jurisdictions have developed processes that provide access to data within four months, the target provided to the Panel, others can take a year or longer. The reasons for delays are multifold, such as concerns over data quality, lack of a roadmap on how to access data, limited budgets for supporting research, fear of potential legal liabilities in the case of data breaches, or broader fears that the research may generate embarrassing results (e.g., evidence of poor performance).

The Panel found that legal definitions and interpretations differ across provinces/territories and countries, which can lead to confusion or overly cautious interpretations of whether data can be accessed or shared. As a result, careful ethical judgments must be taken sometimes in the absence of specific laws. However, good governance ensures that data can be accessed while respecting ethical principles and the law. In searching for models of good governance, the Panel found that successful entities in Canada and abroad have developed systems of governance incorporating four cross-cutting aspects — namely privacy, information, research, and network governance — that achieve this goal. The Panel has identified specific "best practices" within these aspects of governance that can provide the necessary guidance to help transform what is known as a culture of caution to a culture of trust.

The Panel concluded that, although access to health and health-related data vary across Canada, the exemplary practices identified in this report clearly indicate the feasibility of an elevated standard of appropriate data access for *bona fide* public interest research.

Table of Contents

List of Figures

List of Tables

1

Introduction

- Charge to the Panel
- The Panel's Approach
- Structure of the Report

1 Introduction

Canadians care deeply about their country's health-care sector, and look to their governments and health-care providers to ensure that it remains healthy and vibrant. For this to be achievable, however, those who are charged with stewardship of this sector must continuously search for ways to improve it and the health outcomes it generates, including more effective interventions and other ways to improve patient care, better ways to address population health, more efficient models for health-care delivery, and innovative approaches to overall system design. All of these potential areas of improvements are fundamentally dependent on research, and that research is fundamentally dependent on data.[3]

Much of the data relevant to health research arise simply from the myriad of interactions between individuals and the various parts of the health system. Every encounter with a physician, a pharmacist, a laboratory technician, or hospital staff generates data, and much health research relies on gaining access to the records of such encounters. Increasingly, other forms of data are also being generated within the health-care sector, for example by imaging technologies and genetic testing, and research can increasingly take advantage of those new forms of data. Finally, there is a long-standing appreciation that health research needs to include other kinds of data in addition to those from health-care encounters, ranging from risk factors like obesity and physical activity to social and economic backgrounds, which in turn require access to social data from census results, survey data, and administrative data collected by others.

The current state of affairs in health research and its dependency on data can be summed up as follows. First, there has been an explosion in the sheer volume and variety of data generated by and available to the health-care sector and those responsible for Canadians' health. Second, advances in information technology (e.g., lower costs of data storage) are helpful; however, to leverage these data for research and innovation, significant investment in data integration, data linkage, and data analysis is essential. Finally, such data constitute an extremely valuable resource in meeting the challenges of improving health outcomes, managing costs, and accelerating health-care sector innovation.

3 The Panel uses "research" and "researchers" in this report as generic terms to describe analysis of data and those analyzing data, regardless of whether they are in academia or government.

As the following quotation from Neelie Kroes, former Vice-President of the European Commission responsible for the Digital Agenda, indicates, there are still important challenges in gaining access to and using the health and health-related data that are becoming so critical to research and health system innovation:

> Because research in genomics, pharmacology or the fight against cancer increasingly depends on the availability and sophisticated analysis of large datasets, sharing such data means researchers can collaborate, compare, and creatively explore whole new realms. We cannot afford for access to scientific knowledge to become a luxury, and the results of publicly funded research in particular should be spread as widely as possible.
>
> (Kroes, 2011)

In short, data are only useful if they can be accessed and shared. Their value increases when they can be compared with or "linked to" other data collected elsewhere for other purposes. This presents two sorts of challenges.

First, if data have been recorded using different definitions and different methodologies, the process of integrating or comparing databases for joint analysis becomes more complex and may require new analytical tools.

Second, depending on the protocol being followed, connecting different databases (each containing data about an individual's encounters with the health-care or social service system, or their response to a survey) may require that *someone* has access to information that could identify the individual involved. This creates potential risks to the confidentiality of the individual's data, even if the person with access privileges has no interest in the identity of the individual.

Fortunately, access to health data does not occur in a vacuum, but rather within an extensive infrastructure of organizations, each with its own arrangements for the security of computers, networks, and physical records, and each with its own policies and guidelines on how and when access is to be provided. The key challenge then is to ensure that the overarching principles governing the policies and guidelines of these organizations accomplish both fundamental goals involved: enabling access to health and health-related data for sound and ethically robust research while protecting confidentiality of personal information.

There are many success stories in Canada and in other countries where organizations have developed policies and practices allowing appropriate and timely access to health data while at the same time managing potential risks to confidentiality. This report attempts to capture the lessons that might

collectively be learned from these success stories. Of these, maintaining public trust in how data are accessed and used is critical. Fostering greater public understanding of the benefits of research, consulting patients on design and evaluation of health research, and communicating the steps taken to protect their data play a role in maintaining public trust.

1.1 CHARGE TO THE PANEL

In 2013, the Canadian Institutes of Health Research (CIHR; the Sponsor) asked the Council of Canadian Academies (the Council) to respond to the following charge:

> *What is the current state of knowledge surrounding timely access to health and social data for health research and health system innovation in Canada?*

The charge also included five sub-questions:

- *What is known about how to address technological and methodological challenges (such as variable data quality and comparability) associated with linkage of health and social data from various sources and across jurisdictions?*
- *What is known about the benefits, risks and barriers to timely access to health and social data for health research and health system innovation in Canada?*
- *What are the ethical, legal, and social implications of timely access to such data?*
- *What are best practices for improving access to such data for researchers while ensuring appropriate privacy safeguards and also taking full advantage of the digital data revolution?*
- *What are best practices in Canada and internationally for governance frameworks that facilitate access to such data and maintain public trust in the research enterprise?*

To address these questions, the Council formed the Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation (the Panel). The Panel was made up of 14 experts from the health-care sector, academia, and industry. Panel members had experience as data custodians, researchers, managers of health research organizations, or in legal aspects of health research. The Panel met four times over the course of 2014 to review evidence, deliberate, and formulate its findings.

At the outset, the Sponsor gave further direction on interpreting and refining the charge. First, the Sponsor defined *timely access* as access granted within four months of submitting a data request to an organization responsible for providing the data. Second, the assessment should concern only public interest research (i.e., research conducted by public bodies and/or supported by public funds). Thus, health and health-related data used by private, commercial companies were excluded. Third, the assessment should identify best practices in Canada and other countries for timely access to data that can be linked and integrated for research purposes, rather than explore barriers to accessing data in general. Finally, the Panel's work should encompass all types of health data related to publicly funded research, ranging from administrative health data to genomic data.

Within this context, the Panel chose to use the Health Indicator Framework developed by the Canadian Institute for Health Information (CIHI) and Statistics Canada to define *health data* as data on health status of individuals (e.g., well-being, health conditions), health system performance (e.g., accessibility, effectiveness), and community and health system characteristics (e.g., resources) (CIHI, 2014h). The Panel understood the inclusion of *social data* as non-health data that could influence health outcomes, and hence uses the term *health-related data*. This term encompasses non-medical determinants of health such as health behaviours, living and working conditions, personal resources, and environmental factors (CIHI, 2014h). Although many of the challenges and examples discussed throughout the report pertain to health data, the Panel acknowledges that access to social data may also be challenging, and that social data have proven critical for providing an overall picture of health (for an example, see Box 3.3).

The Panel did not intend to provide a comprehensive diagnosis of the state of access to health data in Canada (for a report with this objective, see Meagher and McGrail (2013)); instead, it focused on solutions and good practices. The Panel found many examples of *good practice*, including approaches for dealing with legal and ethical considerations. These are highlighted in Chapter 4. However, in accordance with the charge, the Panel only identified *best practice* [4] related to governance that could be put in place to enable access to health

4 For the purposes of this report, *best practice* is defined as policies and practices currently in use by entities that collect, analyze, provide access to, and regulate laws surrounding access to, data, that — according to evidence identified by the Panel — are already helping to improve timely access while still protecting privacy. In contrast, *good practice* refers to policies and practices that — based on a combination of anecdotal evidence, literature review, and Panel analysis — have the potential to improve timely access while still protecting privacy.

data. It took such governance to include policies and practices for both data custodians and researchers, among others. Best practices for governance are discussed in Chapter 5.

1.2 THE PANEL'S APPROACH

To address the charge, the Panel reviewed evidence on how institutions provide access to data as described in online descriptions and in journal articles. In particular, the Panel looked at institutions that it felt had enabled access particularly well, as well as some that had not. The Panel also drew on *Strengthening Health Information Infrastructure for Health Care Quality Governance*, an OECD (Organisation for Economic Co-operation and Development) report (OECD, 2013d). The Panel benefitted from presentations and discussions at the 2nd International Health Data Linkage Conference held in Vancouver on April 28–30, 2014.

As it examined institutional practices, the Panel noted those that were common to institutions that had enabled access to data and how they differed from practices used by institutions that had not. Enabling access to researchers outside the data-collecting organization was particularly important, as many countries have developed highly advanced systems of gathering health-related data for clinical care, but their method of enabling access for researchers is less advanced.

The Panel developed a tri-level distinction useful for clarifying the different concepts, challenges, and governance aspects involved in timely access to data (Figure 1.1). This distinction is used throughout the report to highlight the issues and best practices that apply to each level of access.

The Panel also focused on jurisdictions with databases housed at different levels of governments or institutions (such as hospitals), particularly where there were variations in legal frameworks across those levels. Many organizations have developed advanced systems to share information *within* their organizations, but many of the fundamental challenges for researchers based in Canada are related to accessing and sharing information *across* organizations (Level 1) and jurisdictions (Levels 2 and 3). The Panel was especially interested in institutions operating in jurisdictions and countries with publicly funded health care that are not only facing, but have also overcome, challenges similar to those in Canada.

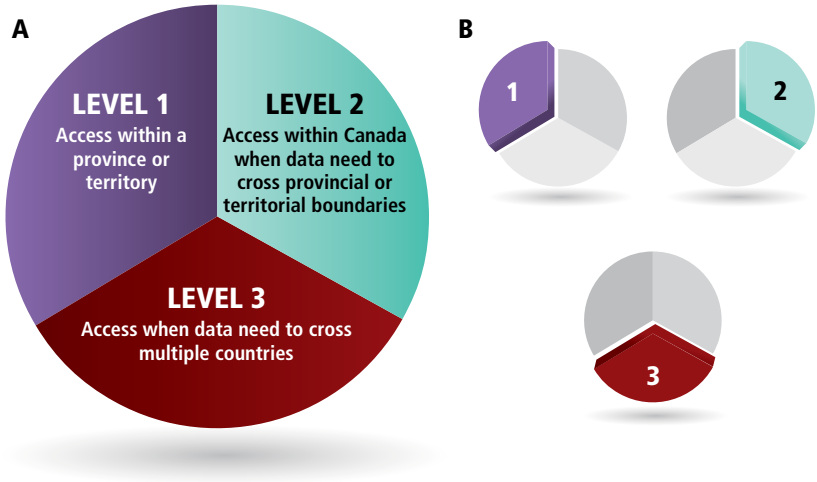


Figure 1.1

Levels of Data Access

(A) Depending on the research question being asked, access to health and/or health-related data may be required from one or more data sets all within one province or territory (Level 1), from data sets spanning multiple provinces or territories (Level 2), or from data sets collected or residing in multiple countries (Level 3). Although more challenges arise when broader access (Levels 2 and 3) is needed, there are challenges related to sharing data across organizations and institutions even within a single jurisdiction. (B) Throughout the report, the Panel uses these graphics to indicate the access levels to which a given concept or issue applies.

From this broad group of institutions and programs, the Panel selected six "best practice entities" (three from Canada and three from other countries) using the following criteria: if not Canadian, the organizations needed to be from jurisdictions with similar legal systems to Canada; they must have developed and implemented governance that has successfully enabled timely access to health and health-related data within a four-month timeframe; and they must have effectively linked databases across organizations within their legally-mandated jurisdiction, all the while affording an appropriate level of privacy protection. The six entities are:

- Farr Institute @ Scotland[5]
- Wales Secure Anonymised Information Linkage Databank (SAIL)
- Data Linkage Western Australia (Data Linkage WA)
- Ontario – Institute for Clinical Evaluative Sciences (ICES)
- Ontario – Better Outcomes Registry and Network (BORN)
- Manitoba Centre for Health Policy (MCHP)

5 The Farr Institute @ Scotland builds on the success of the Scottish Informatics Programme (SHIP), which ran from 2009 to 2013 (SHIP, n.d.-a).

The six entities identified all fit within the first level of access (Figure 1.1) because they deal with data from within a single jurisdiction though, in all cases, from multiple organizations. The Panel found many best practices in the other institutions it examined. For example, Population Data BC provides access to data within one province. Although best practices from single jurisdiction examples were more plentiful, the Panel learned from effective practices at several pan-Canadian (Level 2) entities, including CIHI and Statistics Canada; these lessons are incorporated in the report where appropriate. Most Level 3 examples identified by the Panel, where results have been generated using data from multiple studies across countries, involve genetic data.

The Panel found common priorities underpinning the governance of these entities: enabling appropriate use of data in research, managing risk, protecting privacy, and maintaining public trust. In turn, these priorities are reflected in and inform four cross-cutting aspects of a complete governance approach to access to health and health-related data:

- **Privacy governance:** monitors the risk to privacy posed by data requests from researchers, and the practices of data custodians in providing data (information governance) to ensure that confidentiality is respected. Such governance requires specialized knowledge of technology, law, and statistical methods.[6]
- **Research governance:** ensures that the benefits to society of research outweigh any risks, from both an ethical and legal perspective.
- **Information governance:** determines how organizations and individuals manage information in the health and social services systems to maintain confidentiality of the data while also ensuring appropriate access.
- **Network governance:** involves establishing and publicizing common standards for data collection and developing policies for data sharing within research networks that may span multiple countries.

6 In this report, the Panel distinguished between the concepts of privacy and confidentiality. *Privacy* is a broad concept that in Canadian law encompasses personal *privacy* (protection of one's physical self), territorial privacy (protection of one's private physical space), and *informational privacy* (protection of information about oneself and one's activities) (SCC, 2004). In contrast to privacy, which is a right, confidentiality involves duties. More specifically, *confidentiality* refers to the duties and practices of people and organizations to ensure that individuals' personal information only flows from one entity to another according to legislated or otherwise broadly accepted norms and policies. Maintaining confidentiality in accord with relevant legislation, norms, and policies thereby respects individuals' informational privacy. This report does not touch on any other aspects of privacy. In the context of health data, restrictions on and authorities for communicating personal information arise primarily from legislation, duties relating to professional obligations, or contracts. In these cases, confidentiality is breached whenever personal information is communicated that is not authorized by legislation, professional obligations, or under contractual duties.

Each aspect of governance establishes best practices to ensure that the risk of a loss of confidentiality is managed while fully capturing the benefits of improved health outcomes and appropriate innovation in the health system.

1.3 STRUCTURE OF THE REPORT

The report is organized as follows:

Chapter 2 outlines how researchers access data in Canada and describes how databases can be combined and linked to enable research and analysis, and hence the generation of information that can improve health outcomes and health-care sector efficiency.

Chapter 3 discusses the benefits and risks of linking and accessing health and health-related data, and some key barriers to doing so. It uses examples from Canada to illustrate the health outcomes and health sector enhancements that can be achieved from effective data access and linkage.

Chapter 4 looks at Canada's legal and ethical framework for accessing data for research purposes and identifies good practices in this area. While laws and ethics should guide choices on whether and how to access data, in practice some provinces have succeeded better than others at making their data accessible (Level 1), and differences in legal provisions and interpretations across Canada (Level 2) generally make *bona fide* research access difficult. Research access is even more difficult internationally (Level 3).

Chapter 5 identifies best practices underlying effective governance that defines distinct roles and processes for privacy, research, information, and network governance.

Chapter 6 outlines the report's main conclusions and findings, and synthesizes the evidence and analysis of previous chapters to answer the questions that make up the charge.

2

Data Access, Integration, and Linkage in Canada

- The Explosion in Data
- The Sources of Data
- The Form of the Data
- Using the Data: Transforming Data to Information
- Accessing Data in Canada
- An Example of Data Access in the Netherlands
- Timeliness
- Conclusion

2 Data Access, Integration, and Linkage in Canada

Key Findings

- For effective research with Canada's health and health-related data, many disparate sources of data must be brought together and provided in an "analysis-ready" format, thereby allowing statistical relationships or patterns to be derived.
- Individual-level data held in different databases are more easily compared if they are collected in a standardized manner; otherwise, retrospective harmonization (a second best approach compared with prospective harmonization or standardization) is required to make them comparable.
- Pooled data analysis, which involves physical transfer of individual-level data to a central server, is often hindered by interpretation and implementation of privacy and other laws. Approaches that involve sharing of summary statistics, rather than individual-level data, represent alternative solutions.
- When data are provided to researchers, two different spectra become relevant: an access spectrum (which ranges from secure physical locations, to secure online links, to publicly available websites), and an identifiability spectrum (which involves individual-level data ranging from mildly de-identified to strongly de-identified, followed by aggregated data or data analysis results). As data become more strongly de-identified, the degree of risk to privacy, the access controls, and the scientific value of the data are lowered.
- The ability to access and link data within reasonable timeframes is uneven across Canada or even lacking. Current rules and procedures to authorize research and allow data access overlap and are often time-consuming, processes and requirements for access are sometimes unclear, and access decisions are not always consistent. Delays may be caused by slow approval from research ethics boards, other governing bodies, or data custodians, or by incomplete applications from researchers.

This chapter outlines sources of data that are or could be used for social, health, and health services research; the form in which these data are available; methods for using the data (e.g., pooling, linking, and comparing data from different sources); and current procedures for Canadian-based researchers requesting access to health and health-related data. The digitization of data and proliferation of new monitoring and health-care instruments have led to dramatic growth in individual-level data on factors that affect individual health, social well-being, and the provision of health care. Although routinely collected health-care and social administrative data have been used for health research and system innovation for decades, the rapidly expanding scope of

electronic data provides new opportunities. For example, the sheer scale of big data analytics enables the study of rare diseases, widespread factors with weaker effects, and underserved populations, which may be problematic in conventional research. In addition, the costs of analyzing existing data are substantially lower than those required to perform a new study (Kohane, 2011).

2.1 THE EXPLOSION IN DATA

It is hard to appreciate the tremendous volume of data being created in our current digitized society or to put the pace of its growth in context. In simple numerical terms, the worldwide amount of digital information is expected to grow by a factor of 300 from 2005 to 2020, from 130 exabytes to 40,000 exabytes (or 5,200 gigabytes for each person in the world), doubling every two years. Global information technology (IT) market intelligence firm International Data Corporation (IDC) estimates that up to one-third of these data would be valuable if analyzed, including correlations between scientific data from separate studies and between medical and sociological data (Gantz & Reinsel, 2012).

In the commercial world, many leading firms such as Amazon, Google, and Facebook have made extensive use of data to increase sales and the impacts of advertising. These organizations have an advantage, in that the data are held in their own computer systems and their internet presence is pervasive (Berner *et al.,* 2014; Galbraith, 2014). By contrast, the dispersed nature of the health-care sector — in which data are recorded everywhere from doctors' offices to pharmacies to hospitals — means that data are gathered in many sites. To support high-quality patient care and research, health-care data must be shared across organizations; however, data sharing raises a number of challenging issues (van Panhuis *et al.,* 2014). The challenges are different for social data, where the volume of computerized administrative records (from schools, courts, social assistance, and other sources) is considerably smaller than for health-care data. Social data, however, are often much more varied in form and content, and more dispersed across organizations (Roos *et al.,* 2008; IOM, 2014).

In the future, data provided by self-monitoring from individuals could also be linked to medical records, thanks to greater interconnection between devices and the emerging "Internet of Things"[7] (Simonite, 2013). For example, real-time monitoring of blood pressure and heart rhythm is already being

7 Many electronic devices such as light switches, thermostats and refrigerators are now able to connect and share data via the Internet, and they can be controlled directly via the Internet as well. These devices are broadly defined as the "Internet of Things" (Simonite, 2013). According to Gartner Inc. (2013), there will be nearly 26 billion devices in the Internet of Things by 2020.

fed into medical information systems to provide warning of heart attacks (CHI, 2014b). Private-sector companies are working on a "smart contact lens" to provide a continuous, minimally invasive measurement of the body's glucose levels (Novartis, 2014).

An example of the power and practical applications of health data analysis is a sophisticated study performed by Dr. Douglas Lee and colleagues at the University of Toronto and ICES in Ontario. Using comprehensive provincial health data and validated diagnostic information, they developed a computer algorithm to predict the probability of death within seven days of presentation with acute heart failure at any one of 86 major hospitals in the province. The algorithm can be loaded on a hand-held device and used by staff in emergency departments to inform clinical judgment on which patients require admission to hospital for more intensive management and which patients can be safely discharged. With universal health coverage, the health data for a study like this one are comprehensive, capturing deaths and other major clinical outcomes occurring anywhere in the province and in all provincial institutions. In the future, this type of research will be improved by linkage to electronic medical records (EMRs) and laboratory data, and by the increasing availability of genomic and other sophisticated bioinformatics data (Bodnar, 2012; Lee *et al.,* 2012). This example illustrates the potential for the novel use of data for health-care innovation.

Another major avenue for the use of such data is for research into patterns of diseases or of health-care service utilization. In these cases, the focus is not on the care of particular individuals, but on correlations and other statistical relationships that become apparent only from data spanning large numbers of individuals. Identification of these associations is valuable for directing future inquiries that may reveal causal relationships. An important example is adverse drug reactions (ADRs). ADRs that occur rarely may not be identified or linked to use of a specific drug unless records of many individuals receiving the drug can be analyzed (Senate, 2014).

Similarly, in social science research a large amount of individual-level data is needed to undertake statistical analyses, but individual identities are irrelevant. For example, the Statistics Canada demographic projection model (see Box 3.7) involves access to a wide range of administrative data (on immigrants, births and deaths, and other demographic factors), as well as census and survey data, to produce widely used projections of population characteristics across Canada in future years (StatCan, 2010a).

Despite the explosion of electronic systems and records in all aspects of society, health-care data are not yet fully digital. In 2014, 77% of family physicians used EMRs, up from 41% in 2010 and 64% in 2013 (CHI, 2014a; NPS, 2014). As greater use is made of these new kinds of computerized data, major contributions to research and health system innovation can be expected. Chapter 3 discusses the potential benefits of such data for different fields of research (e.g., epidemiology and genetics). However, it should be noted that adoption of EMRs by physicians does not guarantee their availability or usability for research (Rose *et al.,* 2005). Although initiatives to collect and analyze data from individual EMRs across Canada are under way (e.g., the Canadian Primary Care Sentinel Surveillance Network (CPCSSN, 2013)), in some provinces there is currently no clear framework governing the flow of data from a physician's EMR to the system-wide electronic health record (EHR) (OMA, 2013). In addition, differences in the design of EMR software by competing vendors lead to challenges in interpreting and integrating data collected by physicians using different systems (OMA, 2013).

Whether used for research, quality improvement, health system management, or innovation, timely access to data is essential. Lengthy delays in accessing data can affect the efficiency of research and relevance of findings, and, ultimately, result in lost opportunities to realize benefits that are in the public interest. Further, data that have already been collected for one purpose, such as patient care, provide greater value if they can be used for multiple purposes and in different ways. Clinical practice and policy innovations are ideally based on evidence, sound performance measurement, variance analysis, and management of these measures and variances towards a defined clinical end. The data infrastructure required for research, analytics, and innovation is for practical purposes the same, although there will be differences in the ways the data are used (NSS, n.d.-b).

2.2 THE SOURCES OF DATA

The focus of the Panel's work is primarily individual-level data because timely access for research has been the greatest challenge for these data, and they offer great potential for research and health sector innovation. The vast majority of current individual-level data are created for reasons other than research, such as direct patient care or other types of service provision, which in turn involves transaction processing (for example, writing a prescription, ordering and retrieving a laboratory test or x-ray image for a specific patient) and "narrow" administrative uses (such as billing by physicians remunerated on a

fee-for-service basis). The growing use of EMRs and EHRs has tremendous potential to enhance the quality of computerized data for health research. These electronic records provide an opportunity to close a major data gap in primary health care and other settings by improving the timeliness, quality, and completeness of data, while reducing the burden of data collection and structuring.

Significant progress has been achieved across Canada in the adoption of EMRs and EHRs; this progress has been driven primarily by the needs of patient care. However, the companion software systems, governance, funding, and analytical capacity needed to support research and statistical data analysis from this rich array of digitized patient care data vary across the country. While Canadian entities such as ICES in Ontario and MCHP in Manitoba are world-renowned, in other cases practice in Canada significantly lags behind best practice in other countries.

Data are generated at every encounter between an individual and health-care providers: doctors, hospitals, pharmacies, laboratories, and home-care providers. Figure 2.1 shows the many players involved when moving from collecting primary data (on the left) to combining and integrating data (on the right) to provide the richly detailed and large-sample data sets that hold the greatest promise for contemporary research and analysis. Current programs that pool these data at the health region and health ministry levels, and nationally at CIHI, mainly support health system management and health policy. In some cases, data also flow to provincial health quality councils and university-based research centres, such as MCHP, Population Data BC, and ICES. These research centres (at Level 1) are among the most sophisticated and prolific producers of health services and related research in Canada.

From a broader health perspective, health-care encounter data are increasingly being combined with health and health-related survey data. This was pioneered by Statistics Canada, as well as a number of within-province efforts, and more recently was joined by the pan-Canadian plans of large-scale research platforms such as the Canadian Longitudinal Study of Aging (CLSA) and the Canadian Partnership for Tomorrow Project (CPTP), discussed in this section. Linkage of data across provinces (Level 2) remains a major challenge for these platforms. In social sciences, other administrative data sets are being combined to improve timeliness and accuracy of information; for example, Statistics Canada combines household surveys and university student profiles with income tax records under the auspices of the *Statistics Act* (Finnie *et al.,* 2014).

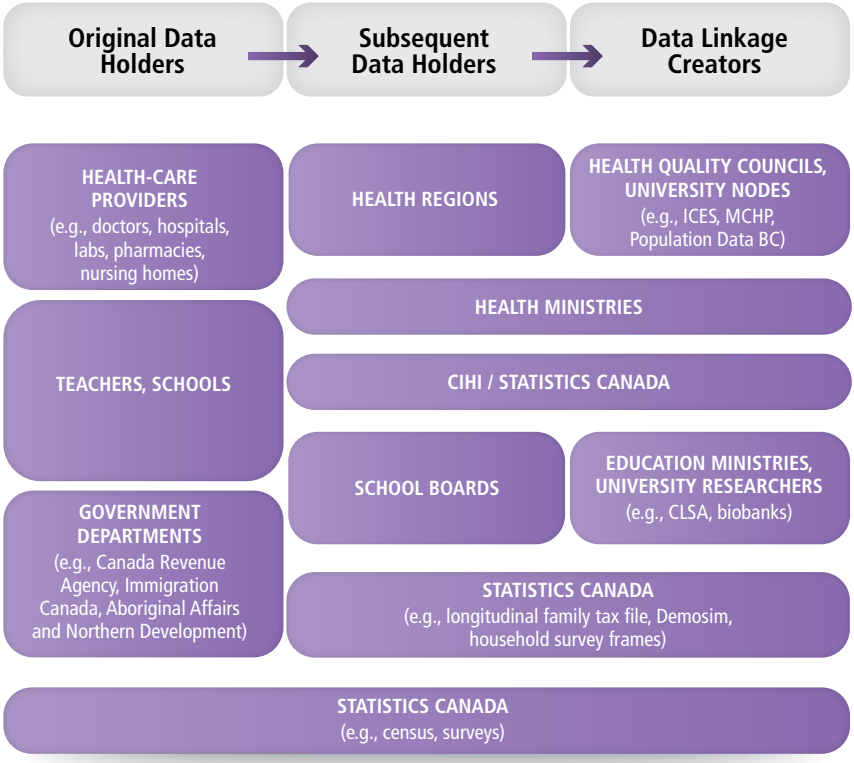


Figure 2.1

Creation and Flow of Data in Canada

This figure shows the different entities that are involved when primary data move from their original holders to subsequent holders to entities that are able to integrate data from multiple sources and provide them to researchers. For example, in Manitoba, prescription drug data from all pharmacies flow to an online database (the Drug Program Information Network) maintained by Manitoba Health. The MCHP at the University of Manitoba houses these data in a large repository, which includes health, social, education, justice, and other data. Researchers may apply to access these data through the MCHP.

Increasingly, administrative data are also being reused. Reuse of data not only lowers the burden on the population and the health-care sector of responding to surveys and filling in forms. It also substantially eliminates issues of selection bias in analysis because the entire population (for example, all births or all cases of cancer) is often captured in these data. This enables valid statistical analysis, thereby fostering research breakthroughs (Jutte *et al.*, 2011). Other benefits of re-using data are discussed in Section 3.2.

Primary encounter data are held, in the first instance, by a *data custodian*. Data custodians collect data and make initial decisions on data use, disclosure, retention, and disposal. They play a central role in enabling or inhibiting access to health and health-related data by implementing policies on data collection, use, and disclosure. They also endeavour to ensure that their employees follow appropriate practices, such as keeping data secure (Chapman, 2005; Province of B.C. & DataBC, 2014). Data custodians include ministries of health, hospitals, physicians, and regional health authorities (Cavoukian, 2005).

In addition to data on health encounters, parallel data sets are being collected and pooled or linked in other domains. Social service encounters, interpreted broadly, include those concerning housing or disability support, educational attainment and performance, and immigration. For example, data on students (at all levels of formal education) are typically generated by their teachers and various kinds of standardized tests. These data flow initially to school boards, and then may be used at provincial education ministries, and then at the national level by the Council of Ministers of Education. These education data are now starting to be used for analysis with health-care encounter data (Brownell *et al.,* 2006; Roos *et al.,* 2006). An example of this is given in Box 3.3.

Other potentially important data from a research perspective are generated by various government ministries, such as immigration data, as a by-product of service delivery. Individual income tax returns from the Canada Revenue Agency now provide the large majority of income data in lieu of income questions on Statistics Canada's surveys, and income data are widely used in both health and social science research.

Researchers are also creating data infrastructures through developing specific longitudinal panels, such as the CPTP and the CLSA. These databases are referred to as *platforms* since they provide the foundation on which many more focused studies can build. The CPTP research platform recently met its initial target, with nearly 300,000 people enrolled. From questionnaire data, biomarker samples, and various physical measures provided by the participants, researchers in Canada and around the world will investigate why cancer and other chronic diseases develop (CPTP, 2011). The CLSA "is a large, national, longitudinal study that will follow approximately 50,000 men and women between the ages of 45 and 85 for at least 20 years. The study will collect information on the changing biological, medical, psychological, social, lifestyle, and economic aspects of people's lives. These factors will be studied to understand how, individually and in combination, they have an impact [on] both maintaining health and

[on] the development of disease and disability as people age" (CLSA, 2009). The challenges of conducting these two nationwide longitudinal cohort studies are discussed in Section 2.5.4 and Box 3.11.

2.3 THE FORM OF THE DATA

An important role of modern IT systems is gathering large amounts of data to use as raw inputs into complex analyses, which transform the raw data to useful information. *Information* is the output of processes that analyze, summarize, interpret and otherwise represent data to convey meaning (DH, 2013). The form in which health and health-related data are available varies greatly at the moment. Some can be "processed, searched, queried, combined, and analyzed relatively straightforwardly," whereas others, typically more qualitative in nature, are harder to combine and analyze (Kitchin, 2014). Together, these data, as well as other types of data that may not seem immediately connected to health, can potentially provide a holistic view of the factors that may be influencing an individual's health and well-being (Weber *et al.,* 2014).

2.3.1 Structured and Unstructured Data

Health and health-related data typically consist of information on various entities (units of analysis) such as individuals, health-care providers, health-care encounters, and social service encounters. In turn, the information on each of these units of analysis consists of a set of attributes. For an individual, these attributes usually include age, sex, and other demographic characteristics and, in the health-care context, symptoms, diagnoses, and interventions. For a health-care encounter, relevant attributes might include time and location and who the providers were. While such use is still nascent, in the future another attribute may be a real-time data stream from a personal health monitoring device, such as the glucose-sensing contact lens mentioned in Section 2.1. Much of these basic data can easily be input in standard formats and handled by automated databases; these are considered *structured data* (Raghupathi & Raghupathi, 2014). Other information (e.g., diagnoses) can be codified using a *controlled vocabulary* such as the World Health Organization (WHO) International Classification of Diseases, 10th revision (ICD-10) (Kohane, 2011).

In contrast to structured data, *unstructured data* (e.g., free-form text, such as written interpretations of x-rays) do not have a common identifiable structure. These data can often be searched as long as they are digital, but they are more difficult to use for computer analyses (Kitchin, 2014). Non-digital, unstructured data, such as the paper notes and forms that still occupy entire walls of cabinets in doctors' offices, can be used for research only with great expense and time commitment to "chart review" (Lapointe *et al.,* 2012). Although there are challenges in reliably extracting meaningful content from free-form digital text,

natural language processing software, which converts textual files into codified terms or tags drawn from controlled vocabularies (Kohane, 2011), has the potential to enable more efficient use of unstructured data within EMRs. Other forms of unstructured data include information generated from the clinical interpretation of diagnostic imaging data and streaming instrument data from wearable biomonitors (MCHP, 2004; Kaminska & New, 2005). In these cases, standardized procedures are needed to generate structured information. This task is routine in some cases: beginning in 2004, Ontario has moved pathology reporting from the traditional narrative structure to reporting based on common standards (CCO, 2012).

Unstructured data can make analysis more time-consuming, but structuring data to be useful has presented challenges. In some settings, there are no implemented standards in electronic records for coding such key data elements as symptoms and diagnoses. Furthermore, clinicians may find it easier — or at least perceive that it is easier — and more informative to use free-form notes (Raghupathi & Raghupathi, 2014). In addition, even where coded data (rather than free-form text) are used in EMRs, in many cases, software vendors have not adopted open standards, such as controlled medical vocabularies (Krist *et al.,* 2014).

2.3.2 Big Data

The term *big data* is becoming increasingly popular (Atkinson, 2014). According to the McKinsey Global Institute, it "refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika *et al.*, 2011). Big data are characterized by the "3 Vs" — volume, variety, and velocity (large amounts of diverse types and formats are generated quickly, possibly in real time) (CHI, 2013). A fourth "V" — veracity — has also been introduced. Veracity is a goal (rather than a characteristic) that refers to the desire for big data itself, as well the outcomes from analyzing it, to be error-free and credible (Raghupathi & Raghupathi, 2014).

Under the above definition and characteristics, certain heath data, such as genomic data and streamed data generated from machines and sensors, could be considered big data. Other health data, such as EMRs, medical imaging and laboratory data, and pharmacy and billing data, have been identified as big data by the literature (Atkinson, 2014; Raghupathi & Raghupathi, 2014; Ross *et al.,* 2014); however, this latter list includes data that are able to be stored and analyzed using typical hardware and software, and thus their categorization as big data is arguable. For the purposes of this report, the Panel also considers administrative data from schools and social services, and large-scale surveys

including the CPTP and CLSA and those conducted by Statistics Canada (especially the Canadian Community Health Surveys), as potential sources of big data.

Regardless of their precise categorization, analytical approaches designed for big data may still be applied to this varied list of health and health-related data (CHI, 2013), particularly if they are used in combination with other big data. In health and health care, the major research potential of big data lies in linking newer forms of data (e.g., genomic data) with more traditional forms to generate and test hypotheses (Raghupathi & Raghupathi, 2014; Weber *et al.,* 2014). Further, by mining (i.e., using increasingly sophisticated statistical methods to analyze) very large data sets, researchers can expand from a hypothesis-driven approach (in which relationships that they expect to see are tested) to a more exploratory approach in which unexpected relationships are uncovered. However, further analysis is generally needed before correlative relationships identified by data mining techniques can be considered causal (Atkinson, 2014). Analysis of big data also often involves processing unstructured data. Whether derived from structured or unstructured data or some combination, new frameworks, technologies, and processes are required to find patterns (CHI, 2013). Applications and challenges of exploring big data are discussed in Section 2.4.4.

2.3.3 Identifiable and De-Identified Data

A simple dichotomy for categorizing the information that is made available for research is to view it as either identifiable or de-identified. As will be discussed later in this chapter, though, in practice there is a spectrum of identifiability. The reason is that there is usually a critical trade-off: the stronger the de-identification of the data, the poorer its value for analysis (AMS, 2006). For example, *de-identification* typically first involves the removal by the data custodian of obvious personal identifiers, such as name and address, from the data file provided to researchers. But depending on what other information is in the data set, further suppression or blurring of the data may be needed to assure de-identification — for example, perturbing birth date, geographic detail, marital status, and the dates of key events like a heart attack. As a result, various degrees of de-identification are coupled with corresponding ease or restrictions on who can access the data, and under which circumstances. Another point to consider is that the level of identifiability — and consequently, the rules surrounding access — may change over time for a given data set. Since the path of a data set cannot be predicted, this point may be taken into account prior to release of data to researchers when risk is assessed (see Chapters 4 and 5).

A central Canadian standard for identifiable information is laid out in the Tri-Council Policy Statement (TCPS), which is the governing ethical standard for Canada's three research granting agencies (known as the Tri-Council) that fund academic research (CIHR *et al.,* 2014).[8] The TCPS states that data are identifiable if they, alone or when combined with other available information, may reasonably be expected to identify an individual (CIHR *et al.,* 2014). There is a lack of clarity about what constitutes "reasonably expected to identify" in practice.

To avoid common confusion about what de-identification means, the Panel has incorporated its concept of reasonableness in the following definition: if data custodians de-identify data, they are undertaking a defensible, repeatable, and auditable process that consistently provides assurance, based on proven and repeatable statistical methodologies, that there is a very small risk of re-identification of any data that are made accessible to researchers (El Emam, 2013b). The Panel has also chosen to use the word *de-identify* rather than other terms in common usage that have various interpretations (such as obfuscate or anonymize). Best practices with respect to this process are outlined in Chapter 5.

There are different rules governing access to de-identified and identifiable data. Depending on the detailed extent of de-identification, de-identified data may be made accessible to researchers pursuant to a spectrum of conditions. In general, there are two main approaches. Under one approach, most often used in smaller academic studies, data are either considered to be de-identified or not. Once they are considered de-identified, researchers have a wider array of options for data access; in some cases, they may be permitted to keep a copy of the data set on their personal computer at home or at their office, and, in others, they may be required to store the data on a secure network (CIHI, 2014b). But for most major research activities involving large volumes of sensitive data, researchers' access to the data involves a combination of de-identification and controlled access. This is the case with data held by ICES, MCHP, and Statistics Canada, for example. Ethics approval by a research ethics board (REB) may be a necessary but by no means a sufficient condition for undertaking research using such data, including virtually all linked health data in Canada (ICES, 2014d; MCHP, 2014d; StatCan, 2014h).

8 The Tri-Council comprises the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council (SSHRC).

2.4 USING THE DATA: TRANSFORMING DATA TO INFORMATION

The Panel's mandate to assess timely access to health and health-related data derives from the potential of these data to support research and innovation. In some cases (e.g., those that require collection of all incident cases of a disease in a population), collecting these data by other means would be virtually impossible or require major new investments. The goal is to enable use of these data to produce innovative information. However, the process of transforming data into useful and accessible information requires both data of high quality and an appropriate infrastructure for managing access to and analysis of the data.[9]

Quality assurance is important for maintaining useful data. Data quality involves dimensions such as accuracy, comparability, usability, relevancy, and whether the data are current at the time they are released to researchers (CIHI, 2014a). The results are documented, problems are noted, and summaries are provided to users so they are aware of issues.[10] Management of data can be facilitated by expert networks that coordinate data dissemination, access, and analysis, such as the Canadian Network for Observational Drug Effect Studies (CNODES), discussed in in Box 3.6.

In many instances, individual data sets are not sufficient to answer the questions addressed; instead, investigators require access to combined data sets. The remainder of this section reviews processes or technical aspects for enabling researchers to make use of multiple data sets (e.g., by comparing or combining information held in different databases). Section 2.5 reviews methods to authorize access. While these sections are presented separately for ease of explanation, in practice there can be complex interactions between the two.

2.4.1 Harmonization to Enable Comparison of Data

The full potential of Canada's health and health-related data can only be realized by bringing different data sets together. These data sets may be from the same province/territory or multiple provinces and territories. An international perspective on health and social issues can, of course, only be acquired by jointly analyzing data from multiple countries. But doing so meaningfully requires that

9 The Panel noted that Canada has an excellent cadre of health researchers; therefore, a sufficient supply of researchers able to analyze these kinds of data is not an issue. However, some learning would be necessary for certain researchers to transition from the analysis of small or medium sized data to much larger data sets.

10 Similar practices exist elsewhere in Canada, such as at MCHP, which has developed macros that automate the data evaluation process. These are available for use and development under a General Publication License (MCHP, 2014g).

the precise meaning of any given data element — whether it is a date, a clinical procedure code, a diagnosis, or a characteristic like "obesity" — is as close as possible across the different data sets being combined (Rabin *et al.,* 2004).

When data elements are *harmonized*, they are considered "inferentially equivalent," and can thus be compared or combined and used meaningfully in statistical analysis (e.g., regression). Harmonization therefore enables data integration, an approach that can be driven by different scientific rationales, including the requirement to obtain sufficient statistical power to investigate relatively rare events or complex interactions. Naturally, when data are collected using identical (i.e., standard) methods and tools, data integration and comparison are facilitated. This is referred to as *prospective harmonization*. However, it can be challenging to develop and implement a set of standards — particularly if underlying consensus is absent — and to maintain standard data collection procedures through time (Fortier *et al.,* 2010, 2012). On a multi-jurisdictional level, this challenge is amplified in Canada since health care is largely provincial rather than federal, thus making it harder to implement pan-Canadian standards (CIHR, 2011). Nonetheless, organizations such as Statistics Canada since its inception (e.g., cause of death coding) and CIHI since 1994 (e.g., hospital procedure and diagnostic codes) have met this mandate (see examples of harmonization below) (CIHI, 2014c).

Even if data were not collected in a standardized/prospectively harmonized manner, valid comparison of information from different studies or databases may still be possible; this approach is referred to as *retrospective harmonization*, and is often critical for making use of existing data (Fortier *et al.*, 2012). Retrospective harmonization could be achieved using rigorous methods, when similar information is collected across data sets. But, the potential to harmonize existing information is necessarily always limited by the heterogeneity of the data collected (Fortier *et al.,* 2011). In addition, certain domains of information are more difficult to harmonize than others. For example, in one study examining the potential for data harmonization, nutritional habit variables were not amenable to harmonization, but almost all variables related to disease history and medication use were compatible with the harmonization tool under investigation (Doiron *et al.,* 2013a).

Examples of Data Harmonization

Prospective harmonization, the ideal approach for data integration, has been achieved in many contexts across Canada, such as acute hospital visits, births, deaths, and cancer information. CIHI plays a central role in supporting data harmonization for health-care encounters, as it is responsible for developing and maintaining health data standards for hospitals, long-term care facilities,

and home-care agencies, among others. Through these standards, CIHI has enabled the production of harmonized data that can be integrated across the provinces and territories. Standards for the collection and interpretation of clinical data are helping to support point-of-care data capture and care delivery. For example, CIHI launched the Continuing Care Reporting System in 2003 to 2004, which uses a standardized clinical assessment instrument to document the characteristics of individuals receiving continuing care (CIHI, 2014d).

To ensure the use of consistent variables and definitions, best practice involves a terminology management strategy, such as that led by the Canadian Partnership Against Cancer on clinical synoptic reporting for surgery (CPAC, n.d.). For such a strategy, the appropriate tools, including thesauruses and glossaries of concepts, need to be developed. These contain, as a minimum for each concept, a standardized term, definition, detailed source information, and related terms.

To process data that have been collected using different standards, retrospective harmonization is required. One such approach is to map one set of clinical diagnostic codes to another — for example, converting ICD codes to Diagnostic and Statistical Manual of Mental Disorders (DSM) codes or vice versa (New Zealand Ministry of Health, 2014). Although Canada currently uses the WHO's ICD-10 codes, research with data from previous years within Canada, or with multiple countries, might require transformation from one version (ICD-9) to the next (ICD-10). In these cases, WHO and member countries collaborate in the construction of "crosswalks" between successive versions of the ICD (AMA, 2012; CIHI, 2014e).

In other contexts, techniques such as DataSHaPER developed by Maelstrom Research (P3G, 2015) have been created to support retrospective harmonization and help determine whether appropriate and valid inferences can be drawn (or not) from variables across different studies (Fortier *et al.,* 2011). The methods and software developed by Maelstrom are currently being used to harmonize and integrate the data collected across major European cohorts (Doiron *et al.,* 2013a) and across the CPTP's five provincial population-based cohorts (with a total of 300,000 enrollees across Canada) (CPTP, 2011; P3G, 2015). Despite the limitations of retrospective harmonization discussed above, generation of such harmonized data structures is one key for leveraging Canadian research innovation through use of existing data.

2.4.2 Data Linkage

Data linkage is the process of bringing together individual-level data from two or more different sources containing data relating to the same individual, family, place, or event (Holman *et al.,* 2008). In the simplest case, these databases

can be linked if a personal health identifier (PHI) — a health card number, for example — is recorded in each of the data sets, allowing different types of data for the same individual to be matched. Other personal identifiers used to link data might be names, addresses, birth dates, or postal codes (CIHR, 2005; CIHR *et al.*, 2014). Because the goal of data linkage is to gather different types of information about one individual, who — for the most part — receives health and health-related services in a single province or territory, it typically requires access to data from one jurisdiction (Level 1 access), but in some cases from more than one organization (e.g., see Box 3.3).

If there is a unique identifier available for virtually all the individuals in the two data sets to be linked, then it is possible to perform an "exact" linkage. But even in this case, there may be errors in the way the identifier has been recorded (e.g., reversed digits), or individuals with missing identifiers. In these situations, it is still possible to perform highly reliable linkages using probabilistic methods.[11] In such a linkage process, even though an exact match may not be possible, two records where the individual characteristics are very similar can be matched. If data sets are strongly de-identified using proper procedures, it will generally not be possible to link them. Thus, if data need to be linked, the simplest way, if possible, is to link prior to de-identification. However, secure linkage technologies have been developed, which allow the linking of de-identified data using cryptographic techniques to protect the fields needed for linkage (El Emam & Arbuckle, 2013).

To reduce the risk of loss of confidentiality during linkage of databases, personal identifiers used to make the linkage are usually stripped from the linked file as soon as the linkage is completed and stored separately (StatCan, 2011). As well, the resulting individual-level databases do not always need to be linked permanently to undertake most research. Instead, the critical feature is that the databases are "linkable" (i.e., that records on a single individual can be brought together when needed for a specific analysis) but the link enabling the fusion of an individual's data from different databases can be broken after the research is completed. Alternatively, the link can be kept completely separate by a trusted third party, especially when it is anticipated that subsequent research projects will need similarly linked data. Decisions on whether to establish a link are made on a case-by-case basis, including consideration of concerns relating to privacy and the potential risks to confidentiality.

11 Probabilistic linking was originally developed and applied at Statistics Canada in the 1960s for cancer research (Marion & Thomas, 2004).

An example in which data linkage can play an important role is for pharmacovigilance, the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problems, which is only feasible through examining data that link an individual's drug use to subsequent adverse health events. Without such linkage, if an individual suffers an adverse reaction to a drug and has to go to hospital, researchers in the health-care system may not be able to identify the association between the drug prescription and the hospital admission because the data on each for the same individual are stored in separate databases (Kirby, 2014).

2.4.3 Data Pooling and Alternatives

Data pooling is the process of bringing multiple data sets together for analysis. It often involves data on individuals with the same or similar content but for different population groups or samples. One key purpose of data pooling is to increase sample size. At present, a main driver of data pooling in health research is population genetics: because there are billions of base pairs in a person's DNA, statistical associations require samples in the millions. Although pooling can occur within or across jurisdictions, it is often most relevant to multi-jurisdictional data access (Levels 2 and 3) since data from different provinces or territories will likely be collected and stored in different databases.

Bona fide pooled data analysis involves physical transfer of individual-level data to a central server, where the data are then analyzed as they would be if they were from the same study (with statistical adjustments if needed). Several genomic initiatives are using this approach (see Section 5.2). In numerous important cases, issues involving interpretation and implementation of privacy and other laws (discussed in Chapters 3 and 4) are hindering the pooling of data from multiple jurisdictions, and are therefore a significant concern of researchers. Pooled analysis may also be impeded by data sets that are too large to be physically shared (Gaye *et al.,* 2014).

Various approaches have been developed to avoid the need to pool individual-level data. One such method is summary data meta-analysis, which involves harmonization of individual-level data across different studies followed by statistical analyses at each study site, and, finally, pooling of the non-confidential summary statistics to obtain an overall result (Gaye *et al.,* 2014). One example that — for the most part — uses this approach is the Sentinel System which is currently being developed by the Food and Drug Administration (FDA), to monitor the safety of drugs and other medicinal products using electronic health data on approximately 100 million people. Mini-Sentinel, a pilot launched in 2009, allows the FDA to query privately held health-care data representing approximately 60 million patients (FDA, 2010; Mini-Sentinel, 2014b). In

this system, identifiable data remain in their existing secure environments. Administrative and clinical information held by each collaborating data partner are standardized using the Common Data Model. The Coordinating Center receives a safety question from the FDA and submits it to each partner. Partners then execute standardized computer programs and share summary results, which are typically aggregated data but may include de-identified individual-level data (Mini-Sentinel, 2014a, 2014b). The Coordinating Center further aggregates each data set it receives and sends its overall findings to the FDA (FDA, 2010). The potential of this approach is illustrated by the sharing of this data infrastructure with other networks, including the National Patient-Centered Clinical Research Network and the National Institutes of Health Distributed Research Network (Curtis *et al.,* 2014).

CNODES is another example of summary data meta-analysis. Similar to the Sentinel System, the CNODES approach allows all data to remain in the province in which they were collected. Information from each database is analyzed *in situ* and only then combined as aggregated and non-confidential data; specifically, the only information that crosses provincial borders is statistical results such as regression coefficients. In contrast to the Sentinel System, CNODES does not use a common data model. Instead, a common analytical protocol is jointly developed and shared, which can be executed in each centre, despite differences in the structures of the databases (Suissa *et al.,* 2012). See Box 3.6 for a detailed explanation of the CNODES methodology.

The European Best Information through Regional Outcomes in Diabetes (EUBIROD) project provides a multi-country (Level 3) example of summary data meta-analysis. The project "aims to build a common European infrastructure for exchange of standardized information about diabetes through connected regional diabetes registers" (Di Iorio *et al.,* 2009). Each data partner uses the same standardized tools and procedures to produce aggregated statistical objects that are sent to the central statistical engine for global analysis. The type of data exchange envisaged by the project is legally viable, as long as all participating countries have fully implemented the EU Data Protection Directive in their national laws, thus guaranteeing an "adequate" level of privacy protection across countries (Di Iorio *et al.,* 2009).

The above approaches may be considered for situations in which answers to research questions are required, but legal issues, costs, and other barriers preclude pooling of individual-level data (Suissa *et al.,* 2012). However, there are circumstances where pooling of data across jurisdictional boundaries improves the analysis, and may even be necessary to avoid invalid conclusions. DataSHIELD, represents an important technical innovation that addresses

these conflicting issues (Wolfson *et al.,* 2010). DataSHIELD is an example of federated data analysis, which mimics pooled data analysis without requiring that individual-level data be moved to a central computer. Under this approach, a central analysis computer sends simple analytic commands to secure servers that are storing individual-level data (Gaye *et al.,* 2014). Simultaneous analyses are carried out on each data computer and summary statistics are returned to the analysis computer. In contrast to summary data meta-analysis, which requires that each partner produces its summary statistics in a timely manner, this method enables real-time analysis while still allowing partners to maintain control of their data (Doiron *et al.,* 2013a). In this respect, DataSHIELD operates in a similar manner to the Sentinel Initiative.

2.4.4 Making Use of Big Data

In traditional data analytics projects, structured data are loaded and processed to produce cross-tabulations or other statistical results, which in turn form the core of a report or research study. Analysis can usually be performed with software installed on a stand-alone system (e.g., a laptop). For big data analytics, however, the volumes of data are so large that processing must be distributed across multiple high-performance computer clusters, which rapidly ingest and process raw data (both structured and unstructured), in near real time as needed. These clusters may use platforms that are available in the cloud (CHI, 2013; Peters & Buntrock, 2014; Raghupathi & Raghupathi, 2014).

Big data analytics allows researchers to reveal patterns in massive volumes of existing data without the need for prior specific hypotheses about where these patterns may exist (e.g., are there any genetic variations that correlate with development of a specific disease?). This "exploratory data analysis" approach contrasts with more conventional data analysis, which is focused on testing one or a few hypotheses (Peters & Buntrock, 2014). More conventional hypothesis testing is, of course, also possible — and indeed statistically more powerful — with big data. In either case, however, there will remain concerns that the statistical results for this observational data analysis need not be revealing valid causal pathways. While the gold standard for establishing causality in health research remains randomized experimentation, such an approach is often completely infeasible. For example, it is not possible to vary experimentally an individual's genetic endowment to see whether he or she develops a disease. As a result, in observational research, reliance is instead placed on having closely similar findings from a range of study populations.[12]

12 The value of purely observational studies is debated in the statistical and epidemiological literature, and there are other important features for high quality analysis, such as properly accounting for confounders. Nevertheless, properly conducted observational analyses are generally very widely accepted scientifically.

Big data analytics combined with data pooling and record linkage has the potential to lower the costs and increase the speed of large-scale studies. For example, Box 2.1 provides an illustration with two types of big data — EHR and genetic data.

Box 2.1
Linking EHRs and DNA Biobanks to Support Genomic Research

Genomic research identifies genetic variations among individuals that are associated with various observable traits (phenotypes), such as development of a disease or disease subtype, or better/worse response to treatment. Before the advent of EHRs and DNA biobanks, genomic studies typically involved recruitment of participants, collection of saliva or blood samples for DNA extraction, and questioning of patients for phenotype information. This model requires funding for biobanking infrastructure, research staff, and processing of samples and data from research participants. A more recent model, which has been referred to as EHR-driven genomic research (EDGR), derives genetic data from samples stored in biobanks (some of which were originally collected for clinical care) and phenotype data from EHRs. Since EDGR involves reuse of pre-existing data, it has the potential to generate results faster with significantly lower costs (Kohane, 2011; Denny, 2012).

A critical aspect of EDGR is the establishment of a link between genetic and phenotypic databases. In the United States, the electronic Medical Records and Genomics (eMERGE) Network involves several medical institutions that maintain biobanks linked to EMRs. The network has published numerous studies that link traits such as body mass index in children, adult height, and risk of hematologic cancer to variations across the genome (Namjou *et al.*, 2013; Schick *et al.*, 2013; Wood *et al.*, 2014).

Studies based on an EDGR model currently face several challenges; one is the development of an acceptable consent regimen. Since samples are often obtained as part of medical care, some research networks and biobanks use an opt-in procedure to recruit patients, whereas others require patients to opt out if they do not want their clinical samples to be used for genomic studies (Kohane, 2011). Another challenge involves developing algorithms that accurately identify cases and controls from phenotypic data in EMRs (Denny, 2012).

2.4.5 Data Management

To cope with the tide of big data and associated analytics, organizations are increasingly formalizing their data management and use practices. This trend results from legal requirements, recognition of the importance of public trust, and opportunities to improve service through the effective use of data. Data management involves policies and procedures covering the full data life cycle (creation, storage, security, archiving, destruction, etc.). In turn, these activities are tied together through data governance: "authority, control, and shared decision-making (planning, monitoring, and enforcement) over the management of data assets" (Mosley, 2008). Data management and governance are becoming central to IT systems as well as key functions of data custodians.

The transition from data to information is also shaped by legal, ethical, and other norms or guidelines that influence data custodians, as well as the effort and cost required to structure data for analysis, as discussed further in Chapter 3. These have led many data custodians to develop a culture of caution in sharing information. This tendency is widespread: the U.K. House of Lords Science and Technology Committee recently noted that, in the context of genomic medicine, the "Department of Health guidance suggests that this domain is affected by 43 relevant pieces of legislation. There were 12 sets of relevant standards and eight professional codes of conduct. What this has bred is a culture of caution, confusion, uncertainty and inconsistency" (House of Lords, 2009). As Willison *et al.* (2011) note, "nobody wants to be on the wrong side of the law, so initial policy interpretations of the legal requirements have tended to err on the side of restricting access."

2.5 ACCESSING DATA IN CANADA

Researchers usually need access to subsets of data relevant to the particular research to be conducted, rather than to the whole data set. Depending on the data and the degree of identifying information needed, a variety of methods are used to enable access. Aggregated data may be made publicly available on a website; individual-level data may be strongly de-identified and sent to the researcher; less strongly de-identified data may be made available to the researcher through secure online links; or individual-level mildly de-identified data may be made available to a researcher who visits a secure physical facility. The weaker the level of de-identification of the data (e.g., the lesser the extent of data blurring and/or data suppression), the tighter the security and controls placed on the use and user of the data, and the more scientifically useful the data that can then be made accessible to the researcher.

Typically, access to more mildly de-identified data is limited to *bona fide* researchers: those who are formally affiliated with an institution of higher learning, who generate new knowledge and understanding using rigorous scientific methods, who intend to publish their research and share their methods, and who conduct research in compliance with ethical and legal requirements as well as recognized good practice (MRC, n.d.). Access to data sets through the Farr Institute @ Scotland (a Level 1 example), for example, may occur via a *safe haven* — a secure computer system that *bona fide* researchers can access physically by an on-site visit or remotely through secure internet connections (ISD Scotland, 2010b) — or through direct transfer. Before being permitted access, researchers must gain the status of an approved researcher, which has five requirements, including demonstration of appropriate Information Governance training and affiliation with an approved organization (ISD Scotland, 2010a). Some organizations in Canada, such as Statistics Canada and ICES, also offer data access through safe havens (El Emam, 2014; ICES, 2014c).

There is no uniform procedure to access data across Canada; indeed, there are large variations both within (Level 1) and across provinces (Level 2), as well as across different kinds of data. This leads to uneven access for *bona fide* researchers and analysts. Sections 2.5.1 to 2.5.5 discuss the different requirements for access in each of the following contexts: an analyst concerned with health-care quality and innovation in a public-sector agency, a university-based researcher working in one of Canada's leading health services research organizations, a university-based social or health science researcher accessing data from Statistics Canada, a researcher belonging to a major health research consortium, and a university-based researcher conducting a small project.

2.5.1 Government and Quasi-Governmental Agencies

Distinguishing features

- Little or no need for formal or external REB ethics approval, since access is governed by legislation, policy, and other means
- No need for external funding or grant applications
- Often full and streamlined access (varies across provinces)
- Generally for health system management or support of health policy rather than for academic research
- Access requirements under legislation may be less onerous than for academic research, although in practice there is often little or no difference

There are a number of sophisticated analytical groups across Canada working inside government or quasi-governmental organizations, all of which fall under Level 1 access. These include the Institut National de Santé Publique du Québec (INSPQ), the Saskatchewan Health Quality Council, the Alberta Ministry of Health, the Newfoundland and Labrador Centre for Health Information, and CIHI, among others. In general, these analysts can access individual-level provincial health-care encounter data that either have been linked or could be linked both longitudinally and across types of encounters (for example, both hospital and physician encounters). Depending on the province's legislation, access for staff of these agencies may be governed differently than for university-based researchers. For example, in Quebec, for the INSPQ to gain access to a new data set requires approval of the Commission d'accès à l'information du Québec for each request, with less demanding provisions for analysis of data on health-care quality in comparison with academic research. Approval by an REB is not required if certain conditions are met.

2.5.2 Academic Centres with Special Relationships to Provincial Ministries of Health

Distinguishing features

- Substantial core-funded infrastructure (programmers, data analysts, physical premises, computers, creation and maintenance of major linked databases, etc.)
- Physically and electronically secure premises; researchers must work on premises or access data remotely on a secure network
- Close working relationships with provincial ministry of health
- Mix of internal work and projects proposed by external researchers
- Substantial cost recovery often required
- Researchers need REB approval and usually require grant funding

Leading-edge data analysis is also concentrated in university-affiliated centres that have been specially organized or supported by their provincial ministries of health precisely for providing access to and/or analyzing provincial health-care encounter data. Population Data BC, MCHP in Manitoba, and ICES in Ontario are Canada's leaders. These centres hold digital copies of virtually the entire collection of routinely collected administrative health-care encounter data in their respective provinces — though this still leaves many omissions, including privately paid care such as dentists and portions of nursing home

costs, in some provinces the actual values of lab tests (e.g., cholesterol levels, not just whether this test was ordered and paid for by the province), and actual diagnoses from physician visits rather than only billing fee codes.

These centres are similar in many ways, but may house data in different forms, use different access procedures, and provide different services. For example, data held by Population Data BC and ICES contain patient identifiers such as names and addresses, but these do not appear in any data sets provided to researchers. Data at MCHP do not include identifiers, but by combining information from several sources, it might be possible to re-identify someone. Population Data BC and ICES have both developed systems that allow researchers to access and process data remotely over secure networks. Researchers can work from their own computers, but data sets remain in protected environments on central servers and only specific research outputs are permitted to leave these environments (ICES, 2014c; PopData BC, 2015b). MCHP data can be accessed at several designated sites throughout Winnipeg (MCHP, 2014d). Scientists associated with MCHP and ICES perform their own research (ICES, 2014e; MCHP, 2014c), whereas Population Data BC acts only as a data provider (PopData BC, 2015c).

Even though specific projects differ, a typical use of the data would start with a *bona fide* researcher developing a research funding proposal involving statistical analysis of specified portions of a group of patients' records — for example to determine how the treatment pathways (e.g., kinds of surgery, extent of chemotherapy and radiotherapy) for breast cancer vary with the patient's age, cancer stage, and hormone receptor status. In this case, ethics approval is part of the process. Further, the cancer stage data may not have been included in the data from the health ministry, so a record linkage with data from the cancer agency may also be required.

Given the level of detail needed for the analysis, the data may still be considered identifiable even though obvious identifiers such as name, address, and health insurance number have been removed. As a result, the computer where the statistical routines are run is in a physically and electronically secure location. A further implication is that the research may be relatively costly. The research grant has to cover both the costs of the data analyst and a fee to allow the centre to recover the costs of its physical space, computers, and such. As the analysis progresses, only tables or statistical results that are clearly non-confidential are allowed to be used outside of the secure facility (CIHR *et al.*, 2014; van Panhuis *et al.*, 2014).

2.5.3 Via Statistics Canada

Distinguishing features

- Legislative mandate to "collect, compile, analyze, abstract and publish" data for the benefit of Canadians
- Large and very detailed individual-level data holdings; wide range of subjects
- Wide spectrum of modes of access for researchers, ranging from access to:
  - aggregate Census and Canadian Community Health Survey data from community profiles on web site
  - public-use microdata (individual-level, strongly de-identified) files
  - researcher submission of statistical analyses over the Internet
  - Research Data Centres in universities and in Ottawa (individual-level, mildly de-identified microdata files)
- Each option along this spectrum involves more stringent rules enabling broader and more detailed access to the individual-level data

As an important Level 2 example, Statistics Canada holds the largest concentration of data for social science and health research in the country. Under the *Statistics Act*, part of the mandate of Statistics Canada includes "to collect, compile, analyze, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and condition of the people" (GOC, 2014a). While a considerable amount of analysis is conducted in-house, Statistics Canada facilitates the use of its data by researchers, provided this research is consistent with its mandate. As a result, there are many ways *bona fide* researchers can access data from Statistics Canada, including the individual-level data that are the focus of this report. From a researcher's perspective, the simplest mode of access is via a public-use microdata file. These are data from a sample survey containing one record per individual respondent. The data have been carefully screened and sufficient data elements suppressed so that Statistics Canada has judged there to be no significant risk that any individual in the data set can be identified (StatCan, 2014f).

Notably, it is a criminal offence to divulge "identifiable" data (Sections 17 and 34 of the *Statistics Act*). In addition, to maintain the trust of Canadians for the volume of data that Statistics Canada collects, it is critical that there never be a data breach (GOC, 2014a).

Statistics Canada recognizes that much of the data it holds would be useless for research purposes if they had to be sufficiently suppressed or blurred to be non-identifiable within the meaning of the *Statistics Act*. For example, the only public-use microdata files are from cross-sectional surveys such as the Canadian Community Health Survey (StatCan, 2014c). Longitudinal surveys, such as the National Population Health Survey and longitudinally linked health-care files that have been provided to Statistics Canada by provinces, are not provided in public-use files because blurring the data sufficiently to render them non-identifiable (in the meaning of the *Statistics Act*) would destroy their value for research (StatCan, 2012).

As a result, Statistics Canada has developed other ways of providing access to *bona fide* researchers while assuring the data remain confidential. One such method is remote data access, where the researcher emails the code for a statistical analysis of a given data set to Statistics Canada where it is run, the output is checked to make sure it does not contain any confidential data, and the results are emailed to the researcher (StatCan, 2014b). Turnaround time with this approach is quite slow.

Another approach is to request access to data at one of Statistics Canada's Research Data Centres (RDCs). More than 26 RDCs are located on university campuses across the country. Researchers first submit a relatively short description of their planned statistical analysis to the funding agency (SSHRC or CIHR), mainly to ensure the proposal meets basic scientific standards (StatCan, 2015). At this stage, the proposal then goes to Statistics Canada, mainly to assess its technical feasibility. Following the review by Statistics Canada, the researcher undergoes a security check, swears an oath to uphold the *Statistics Act*, and is granted access to the data in an RDC as a deemed employee of Statistics Canada (RDC, 2005; StatCan, 2013). RDCs are physically and electronically secure facilities, where researchers are free to apply their statistical analyses to the data. Then, whenever the researcher wishes to take any statistical results out of the RDC, the tables, coefficients, and such must be screened to ensure that no confidential information is divulged (RDC, 2005). While these procedures are onerous, they are deliberately so because sensitive data are made accessible within the RDCs — not only data from longitudinal surveys, but also from the census, tax returns, and health-care encounters, including data that have been linked to health surveys (StatCan, 2014d). Notably, Statistics Netherlands provides an important example in which researchers have access to similarly highly detailed data under more flexible arrangements (see Section 2.6).

Statistics Canada itself conducts many record linkages, among other reasons to improve statistical quality, to reduce respondent burdens, and to economize. Researchers may also apply for custom record linkages. For example, researchers have initiated record linkages connecting air pollution data (based on air quality data by postal code) to hospitalization and mortality data.

Statistics Canada has a detailed policy governing when record linkages are permissible. Essentially, it requires that the analysis to be undertaken with the linked data is in the public interest, and no other means to conduct the research is available at reasonable cost. A description of every linkage is posted on Statistics Canada's website. Any novel linkage requires approval at the most senior level of the organization (StatCan, 2011).

No ethics approval from an REB is required by Statistics Canada for any of these modes of access. In the case of record linkages, senior management approval includes ethics approval. However, the agencies that fund any of these statistical analyses may require formal ethics approval. For example, the Canadian Health Measures Survey, which had substantial funding from Health Canada and included precedent-setting collections of physical measures, underwent a very extensive review by that department's REB, and also produced a very lengthy privacy impact assessment for the Office of the Privacy Commissioner in Canada.

2.5.4 Current Major Health Research Consortia

Distinguishing features

- Massive funding (tens of millions of dollars); multiple sources
- Large groups of researchers (100 or more)
- Very lengthy development periods
- Primary data collections have begun
- Major unresolved issues for pooling linked health-care encounter data

One of the frontiers in health research is bringing together Canada's wealth of health-care encounter data with sample surveys of the population. The main reason is that a host of factors involved in understanding the dynamics of health and disease are not gathered in the routine encounter data — including risk factors like smoking and physical activity; socio-economic background like education, income, and family status; and a growing range of biomarkers from blood pressure to vitamin D to genetics. But such data collections are very large

enterprises, costing in the tens of millions of dollars. As a result, they require a large coalition of researchers to put together the proposals and to develop needed staffing and infrastructure.

At present, there are two main pan-Canadian consortia, the CLSA and CPTP. Their data collections are in the early stages (see Section 2.2). Others that have already been completed (e.g., the Canadian Study of Health and Aging, which examined the epidemiology of dementia in Canada) were smaller in scale. The bulk of the funding for the CLSA and CPTP is devoted to gathering data from survey respondents, including interviews, anthropometry, and blood, urine, and saliva samples. But a crucial planned component is linking these individuals' data to their health-care encounter records where they have given consent. For example, both studies are fundamentally concerned with observing and assessing the likelihoods of developing diseases like cancer in the years following the initial survey data collection to determine reliably the main risk and protective factors — especially factors that are not routinely collected in health-care encounter data.

Both the CLSA and CPTP consortia have begun to work with provincial data custodians (Level 1) to link the health-care encounter records of the respondents in each province to their survey and examination (e.g., biomarker) data. A key unresolved challenge (falling under Level 2) is whether provincial data custodians will allow them to bring the linked data across provincial boundaries so as to achieve the significant and intended benefits of large pan-Canadian samples. Barriers to the CLSA are discussed in Box 3.11.

2.5.5 Other University-Based Researchers

Distinguishing features

- Very heterogeneous
- Mostly smaller projects
- Potential for lengthy, multiple REB approval processes
- Barriers to accessing health-care encounter data
- Negligible capacity/infrastructure to share experiences
- Researchers may need assistance to determine where to find and how to access quality data

In addition to the health research undertaken by university-based researchers via one of the approaches enumerated above, timely access to health and health-care data is central in many typically smaller research projects. One critical step in accessing the data for the research is ethics approval. Since the vast majority of this research is funded by a granting council, prior approval by an REB is required. Once approval is obtained, which itself can be time-consuming, there are further, often time-consuming and frustrating, steps required to gain access to the data.

REBs are expected to give research proposals a thorough review for ethical and legal standards, generally based on the TCPS guidelines. The REB judges whether a project's benefits outweigh harms (including the risk of revealing confidential information about individuals or groups), drawing on ethical guidelines as well as legal requirements (discussed in greater detail in Chapter 3) (CIHR *et al.*, 2014).

To ensure that REBs come to well-informed decisions, the TCPS lays out appropriate competencies that must be reflected in membership of the REB panel, such as having members with specialist legal and ethical backgrounds, as well as representatives of the community. The TCPS does not mandate expertise in privacy practice or law, although such expertise is recommended. Data and privacy are becoming important topics falling within the purview of REBs, but many aspects of technologies related to data do not fall within the traditional knowledge base of all REBs. This absence may have led to concerns that REBs are inconsistent in interpreting ethical and legal guidelines and thus unduly limit the sharing of non-identifiable information (Gershon & Tu, 2008; Willison *et al.*, 2008; Caulfield *et al.*, 2011; Yiannakoulias, 2011; CIHR *et al.*, 2014). These concerns highlight the fact that REBs are now facing an overwhelming list of responsibilities, suggesting that it may be more effective to have dedicated processes in place for evaluation of privacy issues so that the burden does not fall exclusively on REB members without enough expertise in the area. This idea is discussed further in Section 5.2.1.

The REB process evolved in a research context of studies where some intervention was being tested on a sample of patients, mainly at a single centre. Modern research, however, is conducted increasingly in multiple centres and across jurisdictions. Further, the research that is the focus of this Panel does not involve any physical contact with patients at all. Rather, the objective is secondary analysis of data where the identity of the individuals is almost always completely irrelevant. But, complicating matters, the research increasingly requires combining multiple databases to ensure that the sample size is large enough, or when the skills of many researchers are required. As a result, a multiplicity of REB panels may be

involved, and experience has shown that collections of REBs can have inconsistent views. The result too often is significantly repetitious work for researchers and significant delays in the start of research.

An example of the complexity of the REB review process is CARTaGENE (CaG), a longitudinal population cohort of 40,000 citizens of Quebec (aged 40 to 69). It includes both socio-demographic and self-reported health data as well as data from blood and urine tests and anthropometry (CaG, n.d.-c). Participants are asked and typically provide broad consent for future unspecified research. Permission to recontact participants and researcher access to data are subject to approvals from both ethics and access committees.

Today, CaG is well under way, based at the Centre Hospitalier Universitaire Sainte-Justine. CaG is governed by the REB of the hospital and has its own independent Sample and Data Access (SDAC) committee. The access procedure is straightforward, requiring an evaluation of feasibility of the proposed project, submission of required documents (including proof of ethical approval from the applicants' local REB), scientific and ethical reviews by the SDAC and Sainte-Justine, and signing of a Samples and Data Access Agreement (CaG, n.d.-b). In contrast, due to its novel character, the process of initially establishing CaG was complex and time-consuming. See Box 2.2 for a discussion of the challenges involved.

An additional challenge for researchers who are carrying out their own smaller projects is initially identifying which data are available, who they should approach for access to these data, and what the terms of access are. Constantly evolving legal and ethical standards, a multitude of guidance documents from different governing bodies and organizations, and varying interpretations of these documents have all led to confusion and delays (AMS, 2006). To assist researchers, entities such as MCHP maintain a repository of data organized into overarching classes (e.g., health, social, education, and justice data) and further sub-categorized into specific types. For each type (e.g., immunization data), researchers can retrieve a summary that includes, among other details, information about the source agency, the purpose of the data set, the years for which data are available, how the data were collected, and requirements for accessing the data (MCHP, 2014f). This type of guidance is invaluable for researchers who are unfamiliar with the scope of existing data and how to obtain access.

Box 2.2
Setting up CARTaGENE

After five years of consultations, efforts to begin collecting data for CaG began in 2007. The Quebec government insurance board (RAMQ) randomly selected potential participants and, with the permission of the Commission d'accès à l'information du Québec, set up a recruitment calling centre. Since CaG researchers could not contact participants directly, a ministerial decree had to be obtained to allow the RAMQ to be involved in CaG (Awadalla *et al.*, 2013).

CaG is both a population-based biobank and a prospective cohort study (Awadalla *et al.*, 2013). It is an example of a research infrastructure or platform, which means that researchers other than those who initiated the data collection and assembly can also conduct research with these important and innovative data. As a result, it must comply with a multitude of provincial, national, and international legal and ethical guidelines, including (but not limited to) the *Quebec Charter of Human Rights*, the *Civil Code of Quebec*, the TCPS (2010), the Helsinki Declaration of the World Medical Association (revised in 2013), and UNESCO's Universal Declaration on the Human Genome and Human Rights (1997) (CaG, n.d.-a).

While participants had provided consent for research, the ethics approval process became burdensome, ultimately requiring two parallel ethics review systems. The first was a multi-centre approval process with a lead REB that received comments and approvals from 10 local ethics committees. Two did not approve. In addition to seeking approvals from these committees and from the Commission d'accès à l'information du Québec, the Université de Montréal also required ethics approval from its own central REB. As a result, actual recruitment could not begin until the last year of the project. In addition, participants gave consent for access to their RAMQ health-care administrative data, but access to these data is proving to be difficult.

2.6 AN EXAMPLE OF DATA ACCESS IN THE NETHERLANDS

Other countries have developed a variety of approaches to enabling access to highly sensitive individual-level data, including linked data. For example, the U.S. Census Bureau and U.S. National Center for Health Statistics have both created data enclaves on various university campuses similar to Statistics Canada's RDCs. In Denmark, special arrangements are available for *bona fide* researchers to access and analyze their world famous twin registry (SDU, 2014).

A leading example is Statistics Netherlands (CBS), which has, over the past decade, significantly improved the ability of researchers to access detailed highly confidential microdata (anonymous data at the level of individual persons and businesses) for approved research. Staff at Statistics Netherlands' Centre for Policy Related Statistics (CvB) will either carry out custom studies for researchers using these microdata or, under certain conditions, make them available to researchers outside of the CBS. To use the data, researchers must work for organizations that are granted access by the Director General of statistics. Approval by the Central Commission for Statistics is also required in some cases (CBS, 2014b). The data must be used "for the purposes of statistical or academic research" and may only be provided if "appropriate measures have been taken to prevent identification of individual persons, households, companies or institutions from [the] data" (CBS, 2004).

A valuable feature for researchers, which is not allowed under Statistics Canada's current legislation, is the option to access data remotely using a secure internet connection. All data sets used by researchers remain on the secure network environment at CBS and only statistical results are exported for use outside of this environment (CBS, 2014b).

CBS outlines several criteria that must be met by a researcher's institution, so that he or she may be granted access to microdata:

1. "The exclusive purpose of the institution is to conduct scientific or statistical research.

2. If the institution has scientific or statistical research as its primary task, but also performs other tasks or activities, the research component must be effectively separated from these other tasks and activities.

3. The institution must provide public access to its published results, also in the case of research commissioned by a third party: the results of all research for which microdata are used must be made publicly accessible in their entirety immediately and in principle free of charge.

4. The institution must be reputable."

(CBS, 2014a)

As an example, researchers at the Netherlands Interdisciplinary Demographic Institute (NIDI) are studying the impact of the period during World War II when there was a famine in Amsterdam. They have data on individuals whose mothers were pregnant in each trimester during the famine. They are now working with CBS to link these individuals with labour force, hospitalization, and mortality records over the last 25 years to assess the long-term impacts of these extreme prenatal stresses. With NIDI as an "approved organization," there is a secure PC on site with a fingerprint reader. NIDI researchers are themselves approved, as are their specific record linkage and research projects. They are therefore able to access the data and conduct their research at computers only a few steps away from their offices, at any time of day or night.

Based on the criteria quoted above, CBS has approved similar arrangements for researchers at several U.S. and Australian universities, as well as EU universities. CBS has positioned itself as an organization with the potential to become a leading Level 3 example. CBS is also developing means for researchers to use confidential microdata as inputs to researchers' own software such as simulation models.

Clauses to improve access for researchers were added to the *Statistics Netherlands Act* in January 2004 (CBS, 2004). Canada's *Statistics Act* contains does not contain similar clauses (GOC, 2014a). As a result, in Canada it would not be possible under current legislation for Statistics Canada to emulate the Dutch approach for more flexibly providing researchers with access to its data, nor for it to act as a safe and similarly accessible data custodian for other important pan-Canadian data.

2.7 TIMELINESS

A central charge for this assessment was to examine best practice and governance with respect to timely access (see Chapter 1). To this end, the Panel examined available evidence on the time currently taken to access data.

To identify timelines for access to health administrative data across Canada, and factors affecting those timelines, in 2013 researchers at Population Data BC administered a survey of provincial ministries or organizations responsible for access to these data. The study found that when researchers actively pursued an application for data, it took anywhere between 1 to 18 months for them to receive the data. The survey asked organizations what they considered an ideal and realistic length of time for access, and many respondents supported a timeframe of two to four months. Only four (Alberta Health Services, MCHP in Manitoba, ICES in Ontario, and the Dalhousie Population Health

Research Unit in Nova Scotia) of the nine provincial centres participating in the survey reported that they were able to meet this timeframe (Meagher & McGrail, 2013).

In advance of reporting data access times for each organization, Meagher and McGrail (2013) make some important points in their report. They recognize that it is difficult to separate numbers from context, such as variations in the types of data and methods of access offered, as well as "when each centre starts counting time" (Meagher & McGrail, 2013). For example, in all of the centres investigated, research requests required REB approval, with only a few exceptions. However, the approach to data custodian review was more variable, with some centres always requiring individual approval from the originating custodians of each data set, others not including a formal data custodian review but instead considering approval implicit through completion of research agreements, and others setting up access committees to perform the function of data custodians. The first case (individual custodian approval) applied to all projects requiring linked external data. A few centres included additional review processes that were explicitly referred to as privacy reviews (Meagher & McGrail, 2013).

Population Data BC is an example of the first case identified above, in which access is always granted by a designated data steward from each agency that provides data to the organization. Although Population Data BC "acts as the custodian of, and access point for, [its] various data holdings," the holdings themselves "remain under the stewardship of the agency that originally collected the data" (Arabsky *et al.*, 2014).

Arabsky *et al.* (2014) examined the time it takes to prepare data for a data-based research project at Population Data BC, from the time the initial contact is made between the researcher and Population Data BC, to the time that data are delivered. Given that a detailed study on timelines was available for this organization, the Panel chose to use it as an example. However, access to data from Population Data BC takes longer than access to data held at the best practice entities identified by the Panel. The entire process takes an average of just over 15 months (median 10 months). The initial preparation for accessing data (stage one), including developing a linkage approach, clarifying variables requested, and completing forms for ethics approval, takes an average of 4.7 months (median 3 months). Factors affecting the length of this process (referred to as *time drivers*) include the familiarity of the researcher with the application process, which affects how well he or she completes the initial application, as well as the responsiveness of the researcher during back-and-forth communication with Population Data BC.

Obtaining approval from data custodians (stage two) takes an average of 5.6 months (median 4 months), as data custodians review, clarify, and modify applications. During this stage, the data access request must be shown to comply with ethical requirements and privacy legislation. A central time driver at this point is the complexity of the project (e.g., the number of data sources being linked). Population Data BC has made efforts to make stage two more efficient; as a result, the time declined from a median of approximately one year in 2011 to four months in 2013 (Arabsky *et al.*, 2014).

The third stage, which includes completion of agreements related to research, information sharing, and confidentiality, takes an average of four months (median two months). Time drivers at this stage include the responsiveness of all parties involved and the capacity of external data providers to produce the requested data. The final stage of data preparation and delivery takes one month and timing is mainly affected by project complexity (Arabsky *et al.*, 2014).

Access times may be prolonged not only by the complexity of the project, but also by the complexity of the approval process. For example, a request to the Office of the Information and Privacy Commissioner in Quebec requires two levels of approval compared with just one in other jurisdictions (C. Bétie, CAI, personal communication, 2014). Furthermore, the REB process operates separately from review by data custodians. As noted by Arabsky *et al.* (2014), the current process and requirements are not harmonized between data custodians and REBs.

For comparison, the timelines and access rates for the best practice entities selected by the Panel for governance review are reflected in Table 2.1. Similar to the study by Meagher and McGrail (2013), the numbers in the table reflect data provided by the entities themselves and have not been cross-validated by the researcher community, external data sources, or others. They represent the time it takes for a researcher to access data starting from the point at which the researcher submits a completed application and ending when the data are released to the researcher. This is roughly equivalent to the start of stage two in the Population Data BC process, and therefore does not include the time taken for less experienced researchers to submit finalized applications.

Table 2.1
Access Rate and Timelines for the Six "Best Practice Entities" Identified by the Panel

Organization	**Access time** (from point of completed application to point of data delivery)	**Number of requests per year**
MCHP	Approximately 2–4 months	40–50
ICES*	Approximately 1–2 months	Approximately 300
SAIL	Approximately 1–2.5 months (average of 53 days)	Approve 30–50 projects
BORN	Approximately 4–6 weeks	3
Farr Institute @ Scotland	Approximately 1–3 months	Provide access to approximately 150 research project data sets
Data Linkage WA	Approximately 2–3 months	50–100

* Historically, ICES has provided data to researchers appointed as ICES Scientists. Beginning in March 2014, it launched the ICES Data & Analytic Services, which allows any researcher in Ontario to apply for access to data. The numbers in the table reflect the amount of time it takes for ICES Scientists (and not external researchers) to access ICES data.

As described throughout this section, many factors can affect access times, making it difficult to directly compare organizations. (Note that descriptions of the best practice entities chosen by the Panel are provided in Box 5.2.) One critical factor is the scope of information to which the organization provides access. For example, Data Linkage WA offers access to a wide array of both health and non-health data; many requests require data from one or more external providers, such as those that are part of the transport, justice, or social services sectors (D. Rosman, Data Linkage WA, personal communication). Organizations with a more targeted focus (e.g., ICES, health data only; BORN, data related to pregnancy, birth, and childhood) may be in a better position to streamline data access and researchers working in the relevant fields may have the opportunity to become more familiar with data access protocols. Another factor is whether the request involves existing linkages or requires new, custom-made linkages. For example, while SAIL is able to store pre-linked data, the Farr Institute @ Scotland is required to create new linkages for each request (S. Pavis, Farr Institute, personal communication). Thus, the information in Table 2.1 must be considered in the context of these and other variations among the best practice entities.

2.8 CONCLUSION

Tremendous volumes of digitized health and health-related data are generated on a daily basis. These data are created every time individuals interact with health-care providers, educational institutions, and government departments.

These data can take many forms (e.g., free-form text or standard database entries such as age and other demographic characteristics). When the data are provided to researchers, they may be in a form anywhere along a spectrum of identifiability: individual-level data can shift from fully identifiable (made available only under the strictest and most secure conditions) to mildly de-identified to strongly de-identified. The far end of the spectrum does not involve individual-level data; instead, data are aggregated to produce statistical information and only this aggregated information is made accessible.

Canada has health and health-related data with world-class research potential. However, much of this potential is not being fully realized due to an incoherent maze of rules, procedures, and practices, and a general tendency of many data custodians to err on the side of caution when it comes to granting access. Much of this research potential involves secondary use of data — accessing data whose collection is driven, in the first instance, by other very important objectives, including patient care and remuneration of health-care providers. As a result, these data are often not "analysis-ready." They require substantial technical and methodological efforts (e.g., creating derived variables like number of doctor visits in the last year) to make them suitable for statistical analysis. Finally, with health care primarily in provincial or territorial jurisdiction, there is often important variation in nominally similar data collected in different jurisdictions. Therefore, unless they are using data available from CIHI or Statistics Canada, researchers undertaking pan-Canadian analysis face two significant problems: harmonizing data from different provinces and territories, and actually bringing data together across jurisdictional boundaries. Addressing the barriers to pooling data across jurisdictional boundaries would dramatically improve the extent and quality of health research in Canada.

The context of each data custodian and the type of research to be conducted influence the degree of access. As a result, there is no "one-size-fits-all" approach to accessing data; custodians provide and researchers obtain access to data in a variety of ways. As well, long processes to receive access can delay research and engender increased costs, or amount to a *de facto* denial of access. Among the most problematic of these processes are REB approvals, which may involve multiple submissions with differing requirements to many bodies, any one of which may delay or deny approval. Delays may also result from other bodies governing data access (e.g., the Commission d'accès à l'information du Québec) or data custodians. Incomplete applications from researchers can also increase access times. Common approaches that clarify and allocate responsibilities would facilitate and accelerate appropriate access.

3

Accessing Health and Health-Related Data: Benefits, Risks, and Barriers

- Benefits of Health Research
- Benefits of Accessing and Linking Data for Health Research
- How Data Access and Linkage Can Improve Health Outcomes and Health Sector Innovation
- Risks of Research Using Health and Health-Related Data
- Barriers to Accessing Data
- Conclusion

3 Accessing Health and Health-Related Data: Benefits, Risks, and Barriers

Key Findings

- There are major benefits to increasing the appropriate use of individual-level health and health-related data to improve patient care and health and social services.
- Based on evidence from business and government, the main privacy risks of working with identifiable data are deliberate (e.g., malicious attack) or inadvertent (e.g., human error) data release. Although health data breaches can cause serious harm, the risk of a breach actually occurring in the context of research is low, particularly if effective governance mechanisms are in place and if they are respected by care providers, researchers, and data custodians.
- De-identification of individual-level data combined with corresponding regulation of data access is another effective strategy for managing privacy risks.
- An additional risk arises from research on specific communities, which has the potential to make individuals within these communities feel stigmatized. This risk can be appropriately mitigated by involving communities in the research process.
- Numerous barriers can impede health data research, including logistical and ethical barriers to achieving access, reluctance of organizations to share data, and issues with the data themselves, such as lack of comparability between data sets.

This chapter outlines and provides illustrative examples of the main benefits of research based on individual-level data, including analyses of integrated databases and linked health and health-related data. It also reviews the main risks involved and the challenges of using such data. Despite the concerns that an individual could be identified from these rich databases and the often highly sensitive information they contain, the examples demonstrate that important and beneficial research can advance in Canada while confidentiality is maintained.

3.1 BENEFITS OF HEALTH RESEARCH

In 2013, health-care spending in Canada was projected to reach $211 billion — or $5,988 per person — which represents 11.2% of national gross domestic product. Governments' share of health spending has been consistently at around 70% in Canada, with over 90% of government contributions through provincial and territorial governments (CIHI, 2013c). There are widespread fears that, over coming decades, an aging Canadian population and intensification of treatment will increase health-care expenditures (CHSRF, 2011; PBO, 2013).

However, ample evidence shows that better management of health services is key to containing growing health-care costs (Evans *et al.*, 2001). Such management depends critically on being able to work smarter, not harder.

In Canada, economic returns have been estimated for health research projects in general; for instance, a one-time investment of $1 in cardiovascular disease research by public or charitable sources yields a continuing stream of benefits of roughly $0.21 to the Canadian economy each year going forward through better health outcomes (de Oliveira *et al.*, 2013). While there does not seem to be clear evidence on the aggregate contribution of health research in Canada, organizations such as the National Alliance of Provincial Health Research Organizations are demonstrating impact through performance measures, such as the recently developed Canadian Academy of Health Sciences (CAHS) framework to measure returns on investment in health research (CAHS, 2009). This framework has been widely used to measure the impacts of health research in Canada (CCA, 2013).

Regularly collected data, including those capturing individuals' health-care encounters, remain a significantly underutilized resource with the potential to enhance many aspects of health care. Research using these data can improve health outcomes and patient safety, better inform a range of health and social policies, enable beneficial innovations, reduce health-care practices of little or no benefit, and slow the growth in health-care costs (Roos *et al.*, 2008; Taylor & Lynch, 2010; Jutte *et al.*, 2011; Lewis, 2011). However, much of this research is predicated both on timely access to these data and, in some cases, on the ability to link data across jurisdictions. The Panel found many examples of compelling and pertinent research around the world from such data (see Lewis (2011) for a review). It was, however, difficult to quantify the resulting benefits. More information on the scale of these benefits would allow the perceptions and realities of any risks and harms from a potential loss of data confidentiality to be placed in context.

The importance of timely access to health data is effectively demonstrated by examples of the damage that can be done without population-wide collection and prompt analysis of these data. One such example is the story of pemoline, a mild central nervous system stimulant that was approved in 1975 for treatment of children with attention deficit hyperactivity disorder (ADHD). In 1995, researchers at the Hospital for Sick Children in Toronto reported the case of a 14-year-old boy diagnosed with ADHD who had been receiving pemoline for 16 months and methylphenidate for 2 months. The boy died following acute liver failure and an unsuccessful liver transplant. The practitioners who cared

for the boy found two other published cases of acute liver failure associated with pemoline in the United States, but the FDA and the manufacturer were not aware of any others. After the Toronto case was published, additional cases were reported, and pemoline was withdrawn in 1999 and 2005 from the Canadian and U.S. markets, respectively. A 2008 study showed that analysis of existing data could have revealed a significant signal associating pemoline with acute liver failure as early as 1978 (Etwel *et al.*, 2008). Thus, for a period of 25 years, children taking pemoline were unknowingly at risk for a fatal adverse reaction.

3.2 BENEFITS OF ACCESSING AND LINKING DATA FOR HEALTH RESEARCH

There are a variety of well-documented reasons for accessing and linking data sets (Taylor & Lynch, 2010; Jutte *et al.*, 2011). This section outlines some of them:

- **Providing local health benefits:** Improved research benefits policy development and leads to improvements in quality of local health organizations. Medical and scientific knowledge from research improves services and care, and leads to reforms in policy and legislation. For example, in the case of Data Linkage Western Australia, "reforms in mental health legislation and service delivery can be attributed to research based on linked data" (Taylor & Lynch, 2010).
- **Lowering costs and saving time for new research:** Data are often expensive to collect, especially in standardized and computer-accessible forms. Thus, it is cost-effective to take data that were originally collected for administrative or operational reasons and use them for research purposes. Data linkage (see Section 2.4.2) is particularly economical because the costs of developing procedures that make the data linkable are only incurred once (Chamberlayne *et al.*, 1998), thus minimizing costs for new research. Existing information in EMRs also has the potential to simplify expensive, time-consuming clinical trials in certain circumstances, such as those in which approved drugs are evaluated for their comparative effectiveness. Rather than recruiting new participants, researchers can evaluate the outcome of different drugs in patients who are already taking them (Manchester University, 2014).
- **Opening new avenues for research:** Data linkage allows the same data to be used for many different research programs (Holman *et al.*, 2008). Existing population-based data collections can be used more efficiently and effectively than individual longitudinal field studies (Sibthorpe *et al.*, 1995; Brook *et al.*, 2008). Data linked from sources that are seemingly unrelated allow outcomes from different areas (e.g., medical and educational outcomes) to be examined in the same cohort (Jutte *et al.*, 2011). Without linkable data, such research could not be undertaken.

- **Enabling hypothesis generation:** As discussed in Chapter 2, the mining of large data sets and big data can uncover potentially important and previously unconsidered relationships, which can then be investigated further using other approaches, such as retrospective or prospective observational studies, replication in other populations, or, if possible, randomized controlled trials.
- **Improving assessment of overall well-being throughout life:** Linking various health and social data allows researchers to get a more complete picture of the disparate factors that contribute to the physical and mental health of a population. If data collection is ongoing, researchers can evaluate exposure effects over the life span and reach a more sophisticated understanding of temporal relationships (e.g., the impact of conditions during childhood on outcomes later in life) (Hanlon *et al.*, 2007; Jutte *et al.*, 2011). The monitoring, surveillance, and analytical assessment of a wide array of variables within a population may enable knowledge gaps to be addressed.
- **Improving data quality and integrity:** Efforts that promote more extensive use of existing data, such as research involving data linkage, often improve the quality and integrity of these data (Christen & Goiser, 2007). More accurate recording of administrative data may also result from data linkage, as many duplication errors and other technical glitches may be resolved in the linkage process (Holman *et al.*, 2008). The use of data on entire populations — an opportunity for Canada because of its universal health system — would therefore be more accurate.
- **Providing larger, more comprehensive samples:** By enabling researchers to access larger samples, population-wide health and health-related data enable the study of rare events (e.g., rare diseases or rare adverse reactions to treatments), as well as weaker but more pervasive relationships. In jurisdictions with databases that capture information on the entire population (e.g., those with population registries such as Manitoba), this allows the outcomes of individuals who did not receive a service to be included in analyses (Jutte *et al.*, 2011). In some cases, individuals who miss a screening program may be among those most at risk for negative outcomes (Brownell *et al.*, 2011).
- **Reducing bias:** If researchers are able to access existing data without seeking consent from individuals, this can reduce selection, recruitment, and participation biases. These biases, which result from the higher likelihood of certain groups to give consent for access to their records, can potentially lead to inaccurate, misleading results (Tu *et al.*, 2004; Al-Shahi *et al.*, 2005; AMS, 2006; Harris *et al.*, 2008). See Section 4.5.3 for further discussion of this issue.

- **Improving data handling and confidentiality:** Computerized records in appropriate physical, hardware, and software environments are generally much more secure than paper records. For example, automated logging of all access to data is more secure than undocumented access to paper records, which may be held in unsecured environments. Moreover, since data linkage decreases the need for researchers to have access to names and other personal identifiers, it better protects patient confidentiality. Epidemiological and health service research projects using named data from major health collections in Western Australia fell to 36% in 2003 from over 90% before Data Linkage Western Australia was established (Holman *et al.*, 2008). New electronic methods mean that researchers no longer need to find and gather information on individuals manually, a process that was less secure and far more expensive.
- **Lowering response burden:** Filling out forms and answering the same question from different government departments and agencies can be burdensome and costly for citizens. Asking the questions once (in a doctor's office for patient care or an initial study), and then sharing that response between providers and agencies, lowers the response burden for individuals (Jones, 2012). A study in Finland, for example, found that using administrative data is less costly than using dedicated surveys (Gissler & Haukka, 2004).
- **Increasing communication:** Data reuse, including via record linkage, usually results in more communication between researchers, clinicians, administrators, data custodians, consumer groups, and the media (Holman *et al.*, 2008). It "increases team working, allows for improved cooperation and identification of future possibilities and allows for debate about the uses of data and results of any subsequent research" (Taylor & Lynch, 2010).

3.3 HOW DATA ACCESS AND LINKAGE CAN IMPROVE HEALTH OUTCOMES AND HEALTH SECTOR INNOVATION

To illustrate how data access and linkage have been and could be used to enhance health outcomes and health system innovation, the Panel selected some examples of research studies in Canada. These studies have been undertaken by institutions with effective mechanisms to enable timely access to health and health-related data and effectively linkable databases. The research done at these institutions (and at many others in Canada and internationally) is achieved within strong privacy and confidentiality frameworks. Boxes 3.1 to 3.4 provide examples of research using data from within a single province (Level 1), and Boxes 3.5 to 3.8 are examples of pan-Canadian research (Level 2).

Box 3.1 describes research from ICES and Public Health Ontario that used linkage of survey data to show the actions that Ontarians can take, based on sound evidence, to improve their longevity.

Box 3.1
Ontarians Can Live Longer by Living Healthier

Health behaviours play a central role in determining longevity. Researchers at ICES and Public Health Ontario wanted to estimate the risks of death associated with smoking, unhealthy alcohol consumption, poor diet, physical inactivity, and high stress. Information about these risk factors for Ontarians was collected from the Canadian Community Health Surveys completed between 2001 and 2005. Well over 90% of survey respondents agreed to have their responses linked to their provincial health records, which enabled researchers to track mortality rates to 2010. The statistical results were then applied to the 2007 Canadian Community Health Survey, to estimate life expectancy and health-adjusted life expectancy for all Ontarians, and to see how much they changed depending on these five health risk factors.

The results showed that people with the unhealthiest behaviour for all five risks had much shorter life expectancies (68.5 years for men and 71.5 years for women) than people with none of the risks (88.6 for men and 92.5 for women), thereby reducing life expectancy by an average 7.5 years. Three behaviours, in particular, had the greatest association with reduced life expectancy: smoking, physical inactivity, and unhealthy eating. Linked data used in this study demonstrated the real benefits of interventions that can effectively address the most important of these risk factors.

(ICES & PHO, 2012a, 2012b)

Linking data from research cohorts with administrative databases such as medical records can also provide direction for health and social services, as illustrated in Box 3.2.

Box 3.2
Linked Databases Identify Risk and Resilience Factors at One Year of Age

Researchers in Alberta used linked databases to determine what factors protect against adverse outcomes in children at risk of poor developmental outcomes at one year of age. The research used data from a cohort study in Alberta of 3,200 mothers and their children, followed from pregnancy through to school age. The cohort study, called "All Our Babies," was started in 2008 and currently has rich health and social data on the mothers and children over five years. Participating mothers consented to linkage of study data with administrative data from medical records and provided biological samples and infant cord blood as well. Under data sharing agreements, qualified researchers can access the cohort data, subject to research ethics approval of each research project.

Linkage of cohort study data with medical records clearly showed factors that help children at risk because of poverty, language barriers, or mothers with mental health problems or exposure to abuse. Children did well if their caregiver used community supports (such as programs in the community for "Mom and Tot," recreational programs for mothers, or drop-in child care), or played imitation games and read to their children. However, infants born preterm remained at high risk for poor outcomes and required additional supports and services to those described above to alleviate risk of developmental delays. This research provided support for low-cost interventions such as imitation games and reading to children as well as for community programs for mothers and children.

(Gracie *et al.*, 2010; McDonald *et al.*, 2013)

As an example of social research, Box 3.3 shows how record linkage can provide more accurate information about the correlation between socio-economic status and educational achievement than standards test results alone. Assessing student performance can be a useful way to compare education levels across regions and hold education systems accountable for quality education. It can also provide information about educational achievement and socio-economic status.

Box 3.3
Low Income Affects Educational Achievement More Than Previously Thought

While grade 12 testing in Canada has traditionally shown a correlation between low income and educational achievements, these tests consider only students who remain in school until grade 12, and who write the test. Researchers at MCHP used record linkage to analyze data from their Population Health Research Data Repository, including all children who were born and remained in Manitoba until age 18. Linkage between the population registry and the educational enrolment file enabled researchers to identify children and adolescents who were delayed a grade or more or who had withdrawn from school (Brownell *et al.*, 2006; Roos *et al.*, 2006). The analysis showed a stronger relationship between socio-economic status (SES) and achievement than previously found when examining grade 12 standards tests: "only 14% of children whose families had at some point received income assistance passed the grade 12 test on time compared with the 80% of students living in high SES areas" (Roos *et al.*, 2011). The previous approach of looking at the data only for those children who had taken the test had shown that 76% of children from families receiving income assistance passed the test compared with 96% of students living in more affluent areas. Without the data linkage, this considerable discrepancy would not have come to light (Brownell *et al.*, 2006; Roos *et al.*, 2006).

Although many officials in the Manitoba Department of Education had known that there was a relationship between SES and educational performance, when presented with the evidence from the study, government representatives were surprised at the strength of this relationship (Roos *et al.*, 2011).

Box 3.4 demonstrates how data from divergent administrative databases can be linked to provide insight into the health of the population and their use of the health system. The Ministry of Health in British Columbia developed the Health System Matrix to understand the current needs and estimate the future demand for health care in the province.

Box 3.4
Health System Matrix Links Public Health System Data to Understand Health-Care Needs of B.C. Residents

A patient may have multiple encounters with the health system, and information on the encounters is typically collected in separate databases. This presents a significant challenge for both providers and those managing the health system because every encounter may tell only part of the story. The Health System Matrix project in British Columbia creates a summary view of an individual's encounters with the health system over the year by bringing together all major B.C. public health system data sources in a single database. The summary view includes basic socio-demographic data, chronic conditions, and summaries of health-care services that each resident uses. The data are then divided into 13 health status categories (ranging from healthy to end-of-life) and categorized into 25 service lines to show how different population segments use health-care services over time.

The Health System Matrix emphasizes the importance of taking into account the entire health system, which is possible only by combining data sources. The key conclusion from this approach has been that various groups within the population have different health-care needs, use different bundles of services, and move between health states at different rates. This insight has helped to focus the Ministry of Health's strategic planning on prevention, care for people with chronic conditions, and the trajectory to residential care. The Matrix has been used in modelling health care and estimating future demand for it. It has also been incorporated into the population needs-based funding allocation model, one of the tools the Ministry uses to decide on funding for health authorities.

(B.C. Ministry of Health, personal communication, 2014)

Box 3.5 shows how record linkage revealed that seniors are at risk from adverse reactions to drugs. Applying lessons from this study can lead to better health outcomes for seniors, while decreasing the cost associated with unnecessary hospital admissions.

Box 3.5
Seniors are Five Times Likelier to be Hospitalized for Adverse Drug Reactions

Seniors are at an increased risk for adverse drug reactions (ADRs) because of the higher number of drugs they take and the higher prevalence of chronic conditions affecting them. This leads to a higher-than-average rate of hospitalization for people over age 65. In early 2013, CIHI analyzed data for seniors from two of its databases (Discharge Abstract Database or DAD, and Hospital Morbidity Database) in all Canadian provinces and territories over a five-year period. Data from the DAD were linked with drug claims data from the public drug programs in Alberta, Manitoba, and Prince Edward Island — provinces for which linkable data were available. The analysis showed that seniors were five times more likely than the rest of the population to be hospitalized for ADRs. The most likely reasons for hospitalization depended on the type of drug (e.g., blood thinners causing bleeding or opioids causing constipation). Other factors affecting the risk of serious ADRs included maintaining the proper dosage, the total number of drugs taken, drug interactions, patient age, and hospitalization in the previous year.

The linked data revealed serious risks for the first time, and found some of the influences on these risks. The results also put a sharper focus on ways to avoid ADRs in seniors. Medications can be reviewed and managed using drug information systems. A complete picture of an individual's medications can help reduce costly hospital admissions due to ADRs resulting from overmedication.

(CIHI, 2013a)

Box 3.6 shows another approach when data pooling is not possible. Pharmaceutical use is a significant element of Canadian health care. ADRs may be found only after a drug has been approved for use because the small sample sizes typical of the drug trials used for regulatory approval may not provide evidence on infrequent ADRs. As a result, more comprehensive monitoring for ADRs requires data on large populations. A seminal Institute of Medicine (IOM) study in the United States found ADRs to be a major cause of death (IOM, 2000).

Based on studies like that of the IOM, Health Canada recently funded the establishment of a research program to undertake more rigorous post-marketing surveillance. The program assesses possible ADRs by drawing on the linked data sets in a number of provinces. However, because data custodians in the participating provinces were unable to find a means for pooling their data, an alternative approach using "distributed analysis" has been used.

Box 3.6
CNODES Network Uncovers Adverse Drug Reactions across Canada

Until recently, the principal means by which ADRs were detected in Canada was through a voluntary system in which physicians reported side effects to Health Canada.* Confirmation and estimation of risk relied on individual investigators obtaining funding to access and analyze data from individual provinces. This ad hoc arrangement was uncertain and time-consuming. Since 2011, CNODES, created by CIHR and funded by Health Canada, has supported and coordinated a distributed network of provincial teams comprising academic researchers, clinical content experts, and analysts who obtain expedited access to linked data held in provincial repositories.

Because of restrictions on linking health-care data across provinces, the statistical analysis is undertaken in two steps. First, using an agreed and common methodology and analytical protocol, each provincial group analyzes its de-identified individual-level data to see if there is an association between using the drug and a clinical effect (e.g., kidney damage or diabetes). Local sites have been extensively involved in developing the scientific and analytical protocols and have gained experience estimating complex statistical relationships. Aggregate data (regression coefficients) from each province-level analysis are then sent to a central methods group that looks at whether different provincial findings agree and performs a meta-analysis to provide a Canada-wide estimate of effect.

The value of the network has been demonstrated by CNODES studies highlighting the risks of developing kidney damage and diabetes from using high-dose cholesterol-lowering drugs (statins). The studies of statins have involved over two million users identified across seven provinces — something that would be impossible with data from a single province. While the distributed analysis used by CNODES is a significant improvement, there are other approaches. One new approach that achieves the benefits of data pooling without actually requiring data to flow outside any jurisdiction is the DataSHIELD method (discussed in Section 2.4.3).

(Dormuth *et al.*, 2013, 2014)

---

* Vioxx, an anti-inflammatory drug, provides a vivid example in which a clear and serious side effect, heart attack, could not be detected by the voluntary system. The correlation was not easily seen because individual physicians typically have many patients taking Vioxx, of which many have heart attacks. The relationship was seen clearly only after a statistical analysis of millions of patient records in which individual data on drug prescriptions and hospital visits for heart attacks were linked (Graham *et al.*, 2005).

Box 3.7 shows how record linkage can lead to development of a richer picture of the Canadian population, which can provide a more precise context for health policy (StatCan, 2010a). The Canadian population has changed rapidly in recent decades owing to persistent low fertility and strong immigration. Such changes have implications for public policy, which can be better informed by regional population projections concerning variations in visible minority groups, religious affiliation, and mother tongue (StatCan, 2010b).

Box 3.7
Population Trends Can be Projected to Improve Targeted Policies

Public policy can be improved through better information on future demographic trends. In 2004, Statistics Canada developed Demosim, a population projection model that uses a method called *microsimulation*, in which projections are done at the individual level for a representative sample of about seven million Canadians who participated in the former long-form census (and currently its successor, the National Household Survey, in 2011). Demosim also makes extensive use of linked data sets, including census, mortality, income tax, immigration, National Household Survey, and Aboriginal population data — data collected for other primary purposes. The first published Demosim projections indicated that by 2017 more than 50% of Toronto's population would be visible minorities.

The results of the Demosim projections respond directly to policy needs of federal government departments and provide important information to a spectrum of Canadians and Canadian organizations. For example, Demosim is being used to project Canada's Aboriginal population and ethnocultural diversity.

(StatCan, 2010b)

Box 3.8 showcases how data integration, through the use of common data standards and measurement tools, can help improve the health system. Canadian hospital administrators continue to search for ways to optimize the duration of hospital stays while ensuring quality rehabilitation programs and support.

Box 3.8

Secondary Use of Data Leads to Lower Cost and Shorter Length of Stay in Stroke Rehabilitation Units

Health services research to lower costs and improve service delivery to individual patients can benefit from secondary use of data.

In 2001, CIHI developed the National Rehabilitation Reporting System to collect data on patients admitted to rehabilitation programs in hospitals and other centres across Canada. The data gathered include organizational information, patient identifiers, socio-demographic data, administrative data, and clinical information. CIHI then de-identifies and analyzes the data to produce reports for health-care providers across the country.

One such report showed that timely access to care for patients who have suffered a stroke could be improved through better matching of patients to appropriate rehabilitation services. In response, the Calgary Stroke Program developed a new triage system. This triage allowed the program to align patients' needs with the most appropriate rehabilitation services, while discharging low-risk clients to rehabilitate at home with community- and home-based rehabilitation supports. This initiative has resulted in a significant decrease in the cost and average length of stay for individuals in stroke rehabilitation (from an average 72 to 42 days), enabling more patients to have timelier access to the stroke unit.

(CIHI, 2013b)

3.4 RISKS OF RESEARCH USING HEALTH AND HEALTH-RELATED DATA

In the context of health and social data, the conversation between a doctor and patient and the information in a drug prescription are examples of private, personal information that an individual may not want to be revealed. These data could be misused by others outside the health-care context, and lead to stigmatization or discrimination. Hence, in allowing access to data for research purposes, one of the key objectives is to maintain confidentiality of identifiable information. In addition to privacy risks, research on specific communities (e.g., people affected by a particular disease such as HIV) can carry the additional risks of stigmatization or exploitation.

3.4.1 Risks to Privacy

In allowing the use of health and health-related data, one risk is that private information will be revealed. The level of risk to individuals is determined by the probability of revealing the data, and the potential harm from such an event. The Panel identified four main risks (Box 3.9) from using individual-level data, each of which is influenced by the practices of the data custodians and the way in which data are stored and accessed.

Box 3.9
Four Main Privacy Risks of Using Individual-Level Health Data

Risk type 1: Accidental release of data. One risk that attracts much attention is the accidental release of identifiable data — to the public or to unauthorized researchers — when proper security and privacy protocols are not followed, such as forgetting or losing unencrypted USB keys containing large amounts of confidential data in a public place. Any of those with access to data place confidentiality at risk if their data handling procedures are not appropriate.

Risk type 2: Illicit access. Hacking into databases would allow intruders to access data that could be used for illicit purposes. Misuse of data by employees (e.g., employee snooping) may also occur.

Risk type 3: Inadvertent access. Many people may get access to health data in the course of doing their day-to-day jobs. They may inadvertently recognize someone they know in the data set during these activities. Such a spontaneous or inadvertent recognition of someone is considered a breach in the disclosure control community. For example, an employee of the data custodian doing statistical analysis on a data set could inadvertently recognize a neighbour or relative in the database. This is one of the reasons the number of individuals who come into contact with identifiable data needs to be limited, and why proper de-identification needs to be put in place when data will be accessed by many people.

Risk type 4: Data re-identification. Removing information that could identify individuals (de-identification) from data released to researchers greatly lowers the risk of harm. However, if de-identification is done poorly, the data could still contain sufficient information that individuals could be identified and their sensitive information revealed. This is especially true when data are used in conjunction with other public databases and social media. Hence, re-identification risk is the possibility of turning de-identified data back into identifiable data through the use of data matching or similar techniques (ICO, 2012).

Risk types 1 and 2 apply to any type of data (identifiable or de-identified), but are only potentially harmful to individuals if they involve identifiable data; release, loss, or theft of properly de-identified data is not considered a privacy breach. Risk type 3 applies to a small subset of individuals who are authorized to work with identifiable data or with data that are improperly or inadequately de-identified. Risk type 4 primarily applies to data that have not been properly de-identified.

A study by the Ponemon Institute (2013) examining data breaches in 16 industry sectors (including health care, pharmaceuticals, research, and education) suggests that in reality human error (which falls under risk type 1), malicious or criminal attack (risk type 2), and system glitches (which could contribute to types 1, 2, and 3) are the most serious concerns with identifiable data. Shey (2013) reports that, of all sources of data breach, inadvertent misuse by insiders constitutes 36% of cases, and loss or theft of corporate assets (e.g., servers or laptops) accounts for a further 32%. Indeed, concerns over the risks of identifiable data release to unauthorized individuals have become widespread, given media attention to the stories of National Security Agency hacking and thefts of millions of credit card and related data from various large companies. One, however, should be cautious in generalizing from the national security context and a retail context to a research context. Nonetheless, the Panel is aware of a few data breaches of personal health information by researchers in Canada, as well as breaches from research databases (see below). However, there appears to have been no study focusing on the frequency or rarity of such breaches arising specifically within the research domain.

Use of cloud computing for storage of health and health-related data may strengthen data security or add an additional layer of risk. Current systems of governance and accountability are designed with the assumption that data custodians are in full possession of their respective databases and maintain full control over access policies and procedures. In a cloud-based system, health-care and health data organizations are the clients and IT vendors are the providers. When clients entrust their data to systems managed by cloud providers, depending on contractual or other arrangements, they may relinquish some control and may be "unable to exercise technical and managerial measures to ensure access, integrity, confidentiality and transferability of the data" (Seddon & Currie, 2013). For example, providers may have servers in countries or continents outside the jurisdiction from which the data originated, or researchers in one country may access data in another. As discussed in Chapter 4, there is a lack of consensus on which jurisdiction's laws apply (Seddon & Currie, 2013). On the other

hand, *private clouds* (cloud infrastructures set up for single organizations, which govern and control the cloud for their own purposes) can take advantage of many of the benefits of cloud computing, while potentially providing greater security compared with data held on local computers by researchers without an appropriate infrastructure for secure data handling (CHI, 2012).

Examples of Health Data Breaches

Risk type 1 has been influenced by the widespread use of mobile devices such as laptops and USB keys. Two breaches of identifiable health-care data in Canada involved laptop thefts. In one case, a hospital policy requiring data encryption was not followed because the encryption software had not been properly installed. In the other, a laptop containing identifiable data was stolen from a vehicle after it had been removed for research purposes — despite the fact that use of identifiable health information for research was against the hospital's policy. Another Canadian breach resulted from loss of a USB key containing personal health information of individuals who attended flu immunization clinics. An investigation revealed that problems establishing a virtual private network (VPN) had resulted in the use of USB keys to transfer data between eight community clinics (IPCO & CHEO, 2011).

Risk type 1 also includes inappropriate access to identifiable data by researchers, which is more likely if they are working directly from unsecured data sources (e.g., hospital paper charts) without access controls. In this scenario, investigators may or may not be aware that they are violating any research ethics standards. This situation occurred when Canadian university researchers worked on site at a health-care centre to create a registry of health and employment information for individuals in the community working as miners. Although the registry only included information from those who had provided consent, researchers also accessed medical charts from patients who had not consented and created an additional database with de-identified information for statistical purposes. The university maintains that the actions of their researchers did not violate privacy laws or ethical standards, but an investigation by the Office of the Information and Privacy Commissioner concluded that there had been "an improper disclosure and an improper collection of personal information" (Ring, 2011).

Even organizations that house large volumes of health-care data and routinely provide them to researchers can be sources of data breaches if employees do not comply with confidentiality requirements. For example, at one provincial ministry of health, there were three separate incidents involving disclosure of identifiable health information on portable storage devices. In one case, for unknown reasons, the identifiers were not removed from the file; in the other two cases, the employees were not authorized to disclose data to other

employees or external researchers. The disclosures may have been prevented by audit logs or other security measures to detect the access or copying of identifiable information onto unencrypted storage devices (Denham, 2013).

Risk type 2 (illicit access) is a common public concern, but actual breaches rarely involve data that researchers access from secure facilities, in part because they are often de-identified. Hacking of health-care databases with identifiable information may provide a greater incentive. For example, in August 2014, one of the largest hospital networks in the United States reported that the names, addresses, birth dates, phone numbers, and Social Security numbers of 4.5 million patients were stolen from their databases (Weise, 2014). Employee snooping is another type of illicit access, but it is more likely to occur when identifiable information relating to a famous individual is known to be housed at a particular location (e.g., medical records of a celebrity) (Parker-Pope, 2008; Grant, 2014).

Risk type 4 has attracted much attention in the media because high-profile individuals have been re-identified. In one example from 1997, as a proof of principle, a graduate student easily identified the Governor of Massachusetts's medical data within an insurance data set using several pieces of existing knowledge and information from one other database. However, re-identification attempts by data intruders without this knowledge face much stronger challenges; therefore, this example does not support the claim that re-identification is easy and common (Barth-Jones, 2012).

In a systematic review of re-identification attempts on health data, El Emam *et al.* (2011b) showed that most examples of successful re-identification were demonstration attacks performed by researchers to investigate whether a risk was present. Furthermore, most of the data that were successfully re-identified were not de-identified properly in the first place; thus, if existing de-identification standards are adhered to, the risk of re-identification is very low.

All of these examples point to a common theme: the risk of data breach is extremely low if protocols are in place, these protocols are adequate, and they are followed by care providers and their staff, as well as researchers and data custodians. In other words, the level of potential harm resulting from these risks depends on whether there are effective governance mechanisms in place that are being respected.

Another important conclusion from these examples is that breaches rarely occur at institutions with databases set up specifically for maintaining large volumes of health and health-related data for research and administrative purposes; they are

more likely to occur when researchers or employees are accessing data directly from health-care centres. Importantly, there are no examples of breaches at the six "best practice entities" identified by the Panel (see Chapter 1). Risk types 1 to 3 can be limited by good information governance (e.g., secure access to data) and risk type 4 can be contained by proper de-identification techniques.

3.4.2 Risk of Eroding Public Trust

Re-identification of the Massachusetts Governor (discussed above) demonstrated the possibility of correctly identifying a single individual in a database, but only with considerable foreknowledge, which a data intruder may not necessarily possess. Nonetheless, this incident influenced public opinion and the development of privacy policies in the United States (Barth-Jones, 2012). While this successful attack resulted in no gain for the "intruder," it undermined public trust in the processes being used to safeguard privacy.

A significant harm that can result from the erosion of public trust is a change in the behaviour of individuals to protect the privacy of their health information. As reviewed by Malin *et al.* (2013), privacy protective behaviours may involve "going to another doctor, paying out-of-pocket when insured to avoid disclosure, not seeking care to avoid disclosure to an employer, giving inaccurate or incomplete information on medical history, self-treating or self-medicating rather than seeing a provider, or asking a doctor not to write down the health problem or record a less serious or embarrassing condition."

Much of the research investigating patients' perspectives on medical confidentiality has been conducted with so-called "vulnerable populations," such as those with mental health issues, those seeking genetic or HIV testing, and adolescents (Sankar *et al.*, 2003). In one study, a quarter of the adolescents surveyed reported that they would not seek care for health concerns if they thought their parents, friends, or teachers might find out (Cheng *et al.*, 1993). Issues for which adolescents are more likely to withhold information include sexual orientation, drug use, depression, and suicidal thoughts (Lothen-Kline *et al.*, 2003; Sankar *et al.*, 2003). A broader perspective on this issue is offered by Canadian survey data from across the provinces and territories and including individuals aged 16 and older. These data suggest that although 85% of Canadians believe that people withhold health-related information from their doctors, only 28% believe that they do so because of concerns about the security of their health information (Ipsos Reid, 2012).

Patients may attempt to exercise control over their health information at the level of doctor-patient interaction. In turn, physicians may be reluctant to share patient data. In a study involving focus groups with family doctors, reasons for

reluctance included lack of trust in data handling practices of organizations collecting patient data; uncertainty surrounding who the information would be shared with, what it would be used for, and how sharing it would benefit patients; and lack of feedback from public health agencies, which resulted in less motivation to provide data. Physicians felt that data should be de-identified before sharing, and patients should be notified (e.g., through posters in physicians' offices) that their de-identified information might be released to researchers (El Emam *et al.*, 2011c).

3.4.3 Risks of Community Research and Social Sorting

Research using health and health-related data collected from specific communities or groups of people may lead individuals within these groups to feel exploited or stigmatized. However, many of the examples supporting this idea stem from flawed practices. For a sound and ethical study, concerns about the potential for stigmatization need to be balanced with a broader consideration of the potential benefits of the study. Furthermore, as discussed below, collaboration can help to reduce the tension between specific communities and researchers.

There are cases of exploitation of indigenous communities, which resulted from poor research practices lacking in ethical standards, failure to appreciate and respect the culture of the research participants, and failure to collaborate with communities when designing studies (Antone *et al.*, 2014). In two of these instances of exploitation, researchers collected samples from indigenous communities to investigate a particular condition and, without consent, shared them with other researchers and used them for additional purposes (Wiwchar, 2004; Mello & Wolf, 2010).

Genetically isolated communities are often of particular interest to researchers for the study of rare genetic disorders. These communities can benefit significantly from that research. However, there are concerns about how the research is conducted, whether and how research results are shared, whether the results could lead to stigmatization and discrimination, and to what extent the community has access to the benefits of the research (e.g., genetic counselling and testing). For example, in the U.S. case of Greenberg v. Miami Children's Hospital, members of families affected by Canavan disease, who participated in research on the disease and played a substantial role in fundraising and recruiting research subjects, were disturbed that researchers and research institutions had signed restrictive licensing agreements on the genetic tests developed from the research. This meant that they had to pay for access to prenatal testing, which they argued impeded access to the benefits of the research in which they and their family members had participated (Greenfield, 2006).

The above risks of community research can be mitigated by collaborating with community members and organizations representing these communities. This has been demonstrated by ICES, which entered into a data governance agreement (DGA) with the Chiefs of Ontario to protect the interests of First Nations communities (Antone *et al.*, 2014). Because ICES uses de-identified data, it "is able to protect community privacy, thereby respecting First Nations information and governance principles" (Antone *et al.*, 2014). Under the DGA, ICES acts as a steward for First Nations communities, but does not actually own First Nations data (Antone *et al.,* 2014).

Electronic records have supported the process of *social sorting*, which refers to the use of "personal and group data in order to classify people and populations according to varying criteria, to determine who should be targeted for special treatment, suspicion, eligibility, inclusion, access, and so on" (Lyon, 2003). Data linkage can make it easier to collect a wide variety of information on an individual, which could be used to monitor his or her behaviour. For example, the Ontario Works program uses a system of databases to monitor welfare fraud. By linking to government databases, information about a recipient can be collected from the Canada Revenue Agency, Citizenship and Immigration Canada, and Service Canada, among others. Computer applications can use this information along with other data in a recipient's electronic case file to determine how likely an individual is to commit fraud. Various flags may trigger the assessment of a file, such as a recent move or application to post-secondary education. Thus, individuals may feel that they are being unjustly categorized as suspicious (Maki, 2011). This issue is not specific to health and health-related research, but it may be valuable for researchers to consider whether their studies could cause individuals to feel targeted or exploited.

3.5 BARRIERS TO ACCESSING DATA

Despite the potential benefits of using detailed individual-level data, especially linked data, their use is uneven. Barriers to accessing data and linking data sets have been documented (Taylor & Lynch, 2010), and these barriers result in potential delays in accessing data, ranging from three months (Arabsky *et al.*, 2014) to over a year (Meagher & McGrail, 2013), if the data are made available at all. Barriers can also limit potential or interest to use data to generate research and innovation. Several barriers can affect timely access to available data.

Barriers related to access:

- The access process may be unclear for researchers, and they may lack the skills or time to determine how they should proceed (AMS, 2006).
- Easily understandable documentation related to the data sets may be difficult to obtain, but researchers need such documentation to evaluate the usefulness of the data. This includes information on data available (list of variables collected), data format (codes and their meaning), and data collection procedures (population targeted, mode of collection, etc.).
- Access to data can differ depending on the type of users (internal or external).
- There is a lack of resources available for data custodians to generate the data sets required by researchers and to answer their questions (van Panhuis *et al.*, 2014).
- Cost is a factor in enabling timely access to data. Adequate and stable funding is needed to set up a sound infrastructure, attract and retain skilled staff members, and support the continued success of an organization (Marchessault, 2011).
- The many approval processes can cause delays in accessing the data. (For a detailed discussion of REBs and their timelines for approval, see Chapter 2.) In addition to REB review, steps include formulation of the request, review by the data custodian, and delivery of the data (Arabsky *et al.*, 2014).
- To protect confidentiality, the data may be available only through a physical safe haven, but the number of these centres across Canada is relatively limited. Researchers may therefore have to travel to conduct their work.

Barriers resulting from reluctance to share:

- Preparing data to be shared with researchers can be costly; thus, in the context of many competing priorities, data custodians may be reluctant to share if they do not have an adequate budget and/or a specific mandate to support research uses of their data.
- Organizations may fear that their data could be revealed as sub-optimal by others (van Panhuis *et al.*, 2014).
- Anecdotal evidence from Panel members indicates that health professionals and health-care institutions may hesitate to share health and administrative data, including billing data, prescription data, patient safety information, quality assessments, and wait times, which may expose deficiencies that could affect funding or performance evaluation.

- Privacy laws and ethics guidelines or rules may be unclear. Fear of lawsuits may result in incorrect or overly conservative interpretations of legislation (Davies & Collins, 2006), thereby impeding data sharing and other processes that could expedite data access, such as harmonization of REB review.
- The public and health professionals may be afraid to share information due to mistrust of the systems in place to protect data (AMS, 2006; El Emam *et al.*, 2011c). This lack of trust may be influenced by extensive media coverage of data breaches.
- Public bodies and researchers may feel a sense of ownership over their data. For researchers, reluctance to share data may be influenced by competitions for funding, patents or publications, professional recognition among peers, or promotional opportunities (van Panhuis *et al.*, 2014).
- Data custodians may fear that their actions could lead to a data breach, which may damage their reputation or result in public harm; on the other hand, they will likely not personally benefit from providing access. Thus, they face an unbalanced or asymmetric risk when they opt to share data.

Barriers related to data quality, utility, and comparability:

- Data collected for a targeted purpose (e.g., physician billing) are not necessarily ideal to meet the secondary usage objectives of research and innovation. Specifically, the content and structure of the data are not always adaptable or relevant to the intended use for research. (The Panel notes, however, that problems with data sets can only be recognized and remedied if researchers attempt to use them despite their imperfections.)
- Data can require significant processing to be useful for research purposes, which is time-consuming. Institutions need to keep up with rapid changes in data technologies and improve their standards to produce analysis-ready data (van Panhuis *et al.*, 2014).
- Incompatibilities between data management systems across institutions pose a barrier to sharing, linking, or harmonizing data (van Panhuis *et al.*, 2014).
- Heterogeneity of data across pooled or linked data sets can necessitate harmonization (see Chapter 2), which can render the process particularly challenging and time-consuming (Flowers & Ferguson, 2010).

Examples of Other Barriers:

In some cases, barriers arise from the inability to link to certain types of data and difficulties pooling across provinces. An additional barrier involves the implementation of overly restrictive privacy legislation based on assumptions about what the public may find acceptable. These barriers are explored below using several examples. Box 3.10 is a within-province (Level 1) example, and Boxes 3.11 and 3.12 deal with linking and/or pooling data across Canada (Level 2).

Box 3.10 outlines how linkage of detailed health and administrative data sets are required to enable effective health services and system research. Modelling of aspects of the health system is possible in one province because of the existence of linked data sets, whereas it is proving a challenge in another province lacking such data.

Box 3.10
Breast Cancer Screening

There is considerable controversy regarding the age threshold at which women should be offered mammographic screening. Genome Canada and CIHR recently awarded a substantial grant to researchers examining the option of shifting breast cancer screening from being based primarily on age to being based on personal risk factors (GenomeCanada & CIHR, 2012), including conventional factors such as fertility history as well as genetic factors. As part of this project, a specialized computer simulation model is being built. The model uses detailed data on the population distribution of these risk factors to project what would happen if risk-based screening were introduced. The model considers costs of breast screening programs, as well as the number of women currently going through these programs, and then through diagnostic work-up, treatment, follow-up, and disease progression. It also considers the effects of screening and follow-up on women's ability to carry on daily activities such as paid employment.

Adopting risk-based cancer screening will depend, in part, on weighing the costs of screening different populations for each provincial health system against the potential benefit in terms of breast cancer deaths likely to be averted. To make these determinations, linked individual-level data are required to complete the model. Extensive use is being made of Ontario data held by ICES, which, in collaboration with Cancer Care Ontario, has linked detailed cancer incidence data to health-care records. The researchers also wished to use the model for Quebec residents (ICES, 2014f). However, since Quebec has no comparable organization to ICES in Ontario, gaining access to and linking the required data is proving a significant challenge. This project illustrates the benefits of linked health and administrative data in considering a fundamental change to a health service.

Box 3.11 outlines how the CLSA has been built to develop insights into Canada's aging population, a challenge that much of the rest of the world is also now facing. However, as a result of legislative or other barriers, these data have so far not been linked to administrative data or pooled to develop further insights.

Box 3.11
Canadian Longitudinal Study on Aging

The CLSA is a large, long-term cohort study following the health, psychological, social, and economic aspects affecting 50,000 Canadians for a period of 20 years. It aims to understand the impact of these many factors on health, disease and disability as people age. The participants provide a core set of data on demographic, lifestyle, social, and socio-economic factors, among others (CIHR, 2013d).

Those involved in managing the CLSA anticipated that significant benefits would result from linking the original data generated by the study to administrative health-care data (e.g., physician visits, hospitalizations, and prescription drug use). The power of this approach has been demonstrated by the Aging in Manitoba (AIM) Longitudinal Study involving almost 9,000 older Manitobans. The study includes data from several waves of personal interviews collected between 1971 and 2001, which are linked to health services utilization data housed at the MCHP (University of Manitoba, 2014). In general, linkage of population-based administrative data with various external health and social data sets in Manitoba has created an "information-rich" environment for studying health determinants and outcomes (Roos *et al.*, 2004; 2008). Having realized the research opportunities provided by linking administrative health databases and large cohort studies, consent forms for CLSA participants included questions about data linkage (Doiron *et al.*, 2013b). However, a mechanism for such linkage has not yet been implemented, and will require strong coordination and long-term commitment from all stakeholders.

Another significant challenge for the CLSA is that some provinces interpret privacy legislation as not allowing administrative data to cross provincial jurisdictional boundaries. Hence, although legislation would permit the CLSA to link data within each province if the mechanisms were in place, pooling of the linked data would be prohibited. Moreover, some data custodians have indicated that their resources and capacity are too limited to meet data access, extraction, and linkage requests. As a result, data access requests from researchers are generally met only "when, and if, there is time" (Doiron *et al.*, 2013b).

Box 3.12 illustrates how, despite the potential benefits to health services or system research of accessing linked data, there are many instances where the maximum potential of research has not been fulfilled. The issues in this case were legislative and inter-jurisdictional barriers to integrating data, and an inability to link databases due to unavailable identifying information.

Box 3.12
Integrating Data across Canada

By integrating hospital discharge records across Canada, it has long been possible to observe large variations in treatments by geographical area. However, understanding the effects of these unexplained variations on Canadians' health has generally remained difficult. As demonstrated in the unique study below, some important progress has been made by integrating patient records, but the interpretation of these variations remains limited.

Researchers at Statistics Canada (Johansen *et al.*, 2009) used data from the federal agency's Health Person-Oriented Information Database to analyze both treatment and 30-day survival outcomes for heart attack patients from seven Canadian provinces for 1995–1996 and 2003–2004. The results showed a 3:1 variation in the likelihood of having a major health-care intervention for heart attack (heart bypass or angioplasty, with rates between 20% and 60%) with no discernible benefit in terms of survival within 30 days of the heart attack. Moreover, the study showed that some regions with high intervention rates had relatively high mortality rates, while some regions with lower treatment rates also had low mortality rates (Johansen *et al.*, 2009).

These results suggest potentially critical inefficiencies in Canada's health-care system, as well as large variations across the country in heart attack treatment without any obvious explanation. It was very difficult, even with the authority of the *Statistics Act*, to extract and analyze these data. Moreover, the data were still very limited; for example, the analysts were unable to examine longer mortality follow-up (such as mortality rates after one year) because the identifying information needed to link patients to death certificates was not available. To understand why these large variations exist and are growing, more data would be required to help clinicians and researchers determine which types of heart attack treatment would be most appropriate (Johansen *et al.*, 2009). For such longer-term follow-up, accessing these data was impossible because most provinces' officials were reluctant to provide Statistics Canada with identifying patient data (M.C. Wolfson, University of Ottawa, personal communication).

The final barrier discussed in this section is concerned with the potential for lack of public engagement to hinder research. A report by the Academy of Medical Sciences in the United Kingdom on the use of health information for medical research recognizes this issue, stating that a lack of communication with the public about their opinions on the use of personal data for research leads to "defensive and restrictive interpretations of the law, which may not represent the wishes of an informed public" (AMS, 2006).

Although more evidence is available on public attitudes towards the use of health data for care and treatment (rather than research) (AMS, 2006), polling data indicate that 80% of Canadians generally support the use of EHRs for research, with support increasing to 88% if details such as their name and address are hidden from researchers (Ipsos Reid, 2012). Canadians consider the protection of health information to be extremely important, but many are unaware of specific privacy laws that are in place, suggesting unfamiliarity with their rights concerning protection of personal health information (EKOS, 2007; Ipsos Reid, 2012). Informing the public of both the controls that have been implemented to safeguard information, as well as the benefits that have resulted from research using health data, has the potential to increase confidence in and enthusiasm for this type of research (AMS, 2006). However, public engagement is not a trivial task — it requires transparency and receptiveness at multiple stages of the health-care process (Henke *et al.*, 2012), along with sound research to determine how best to involve the public (see Section 5.2.5 for further discussion).

3.6 CONCLUSION

The Panel found evidence of clear benefits to enabling access to health data by researchers. There are many instances of successful research leading to improvements in health care or services, without any significant loss of confidentiality. Ideally, understanding the scale of potential benefits to health care would guide appropriate judgment on regulations and processes facilitating access to health data for research. However, the current lack of evidence describing measurable benefits of research, coupled with public concern over risks of confidentiality loss, can unduly tilt the balance against further enabling access to data. In addition, when working with specific communities, there may be real risks of stigmatization or distrust that investigators need to address in their research approach.

The Panel identified a highly exaggerated perception that data linkage increases the likelihood of disclosure of personal health information. While laws and regulations govern protection of this information, they may be inconsistent. Furthermore, a single regulation is often interpreted inconsistently or too conservatively. These issues have led to uneven data and information governance across Canada and internationally, which prevent access to data that could provide substantial public benefit. Access is also hindered by technical barriers and reluctance to share data. There are numerous reasons for this reluctance. One may be a lack of resources for data custodians, who are faced with many competing priorities. The Panel provided anecdotal evidence that organizations and groups involved in delivering health care may also fear negative publicity if research were to reveal poor performance.

However, there are several solutions to overcome these obstacles. Proper de-identification of data and effective governance mechanisms — which are respected by care providers, researchers, and data custodians — provide effective methods for substantially limiting the risk of identifying people from individual-level data. Public engagement has the potential to increase confidence in and support for research with health data.

4

Accessing Data for Research Purposes: Canada's Current Legal and Ethical Framework

4 Accessing Data for Research Purposes: Canada's Current Legal and Ethical Framework

Key Findings

- There are ethical imperatives to protect the confidentiality of individuals' data, on the one hand, and to provide access to quality data that enable research in the public interest, on the other. These two imperatives need not conflict.
- Data custodians have fundamental legal duties to protect confidentiality of personal data, and these duties underpin their conduct. These duties can lead to cautious and conservative interpretations of allowable access when a complementary mandate to enable access to data for research is not made explicit.
- Canadian federal and provincial/territorial laws generally address identifiable information and do not constrain researchers' access to de-identified or non-identifiable information.
- Given the imprecise and inconsistent definitions of the term *identifiable information* in laws and ethical guidelines from different jurisdictions, it is difficult to be sure whether a data set qualifies as non-identifiable. Instead, it is useful to view de-identification as a continuum and to adjust access controls accordingly to mitigate re-identification risk.
- Canada's governance of research ethics is fragmented, with significant differences across the provinces/territories. As well, laws on sharing data across provinces/territories and between countries differ or are lacking, sometimes leading to confusion for researchers and REBs about whether, or on what basis, data can be shared.
- While participant consent is a cornerstone of experimental research involving humans, the ethical and legal considerations for accessing personal information are not the same as those for physical involvement in research. There may be sound ethical reasons to pursue research with health and health-related data without consent in some circumstances, notably when risk is managed and the research benefits the public good. Appropriate risk management involves keeping measures to protect privacy proportional to the potential harms of proposed research.

Enabling access to data collected about individuals rests on legal and ethical norms. In Canada and in many other countries, respect for privacy and confidentiality of personal health and social information is protected by laws. In addition to the legal framework, ethical principles applied during approval for research use of individual-level data seek to protect privacy, but also recognize the public benefit of research; that is, they recognize the benefits of research to individuals and the public at large, and support measures that facilitate such research.

This chapter looks at the broader framework of the law and ethical guidelines in Canada. This framework guides the balance between access and confidentiality. However, implementing these considerations in practice when data custodians grant researchers access to data may be challenging, as not all implications have been articulated and clarified. Because of the pace of recent technological changes and the rapid growth of information sharing, best practices for many of the legal and ethical issues involved in data sharing are evolving. As well, the law differs among provinces/territories within Canada, as well as internationally, which means that best practices in one context may not be directly applicable in another. Therefore, the Panel did not identify best practices in this chapter, but instead highlighted good practices where possible (see Glossary for definitions of *good* and *best practice*).

4.1 CANADA'S LEGAL FRAMEWORK

This section reviews the legal structures governing the sharing of health information, both internationally and in Canada. The federal government and the provinces and territories have legislation governing the use and sharing of health information for research purposes. Once legal requirements are met, decision-makers also apply ethical guidelines and principles in evaluating research proposals for the use of health information. A detailed compendium of these provisions is found in Table 4.1 at the end of this chapter.

4.2 BASIC PRIVACY PROTECTIONS: INTERNATIONAL LEGAL FRAMEWORKS

Internationally, most industrialized nations have laws protecting personal information or health information. However, there is large variation in the regulations, their objectives, and their restrictions on data sharing across international borders. In Europe, for example, data protection is treated as a fundamental right (EU, 2000), and a common legislative framework that currently influences member states (the Data Protection Directive) has been undergoing revisions. If successful, the reforms will result in a new Data Protection Regulation that will have the direct force of law and be directly applicable in the member states of the European Union. It will serve as the basis for a single, uniform approach across domestic laws in EU countries. By contrast, data confidentiality is not conceived as a right in several other jurisdictions, such as the United States (EU, 2000). The permissibility of cross-border data sharing may depend on the level of legal protection of data in the country providing the data (as applies, for example, to the European Union), or, alternatively, on the measures taken by the individual person or organization that receives the

data (such as in Canada), or on the receiving jurisdiction's ability to realize the benefits of electronic commerce (for example, in the Asia-Pacific Economic Cooperation region) (Kuner, 2013).

4.3 BASIC PRIVACY PROTECTIONS: CANADIAN LEGAL FRAMEWORKS

Domestically, Canadian legislation strives to both protect health information privacy and facilitate information sharing for the purpose of health research. At the national level, key legislation includes the *Statistics Act* (1985), the *Privacy Act (1985)*, and the *Personal Information Protection and Electronic Documents Act* (2000) (PIPEDA). However, most of the relevant legislation is at the provincial level (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GPEI, 2012; GDQ, 2014a, 2014b; GNL, 2014; GOA, 2014b; Govt. of BC, 2014b, 2014c; Govt. of MB, 2014b). As detailed in Table 4.1 and discussed below, these provincial health information laws have differences as well as some common features.

4.3.1 Sharing of Health Data that Have Been De-Identified

In all provinces, "de-identified" health information is not subject to legislative regulation, and data that are not identifiable may generally be shared and used freely by data custodians and researchers (see Table 4.1, row 1). However, there is variation among the provinces in how identifiable is defined: in some provinces, it means information from which a person can reasonably be identified, yet in others, it means data from which a person's identity is "readily ascertainable" (see Table 4.1, row 2). Box 4.1 shows the narrowest legislative definition, from Alberta's law.

Box 4.1
Defining Identifiable in Alberta Legislation

The Alberta *Health Information Act* (GOA, 2014b) defines *individually identifying* health information as when "the identity of the individual who is the subject of the information can be readily ascertained from the information." It goes on to define *non-identifying* to mean that "the identity of the individual who is the subject of the information cannot be readily ascertained from the information." By using this standard of information from which identity can be "readily ascertained," either from the information itself or from the base information combined with other health information, the Alberta legislation has a narrower definition of "identifiable" than other provinces (GOA, 2014b).

Furthermore, there are different definitions flowing from ethical guidelines and court decisions. The Tri-Council Policy Statement (TCPS) governing ethical research in Canada offers a standard definition of *identifiable* that REBs across Canada follow. It states that "information is identifiable if it, alone or in combination with other available information, may reasonably be expected to identify an individual" (CIHR *et al.*, 2014). Among relevant legal cases, in a 2001 Ontario case involving disclosure of the medical procedure charges of the highest billing physician in Toronto, the Ontario Superior Court established that information can be considered personal if "there is a reasonable expectation that, when the information in it is combined with information from sources otherwise available, the individual can be identified. A person is also identifiable from a record where he or she could be identified by those familiar with the particular circumstances or events contained in the record" (OSCJ, 2001). In a 2008 case involving open access to the Canadian Adverse Drug Reaction Information System database, the Federal Court held that "[i]nformation will be about an identifiable individual where there is a serious possibility that an individual could be identified through the use of that information, alone or in combination with other available information" (FCC, 2008).

These variable interpretations of the terms *identifiable* and *de-identified* make it confusing to base data sharing guidelines on the notion that "non-identifiable data" can be used freely. It is difficult to be sure whether a data set qualifies as non-identifiable. In reality, it is only clear that data may be fully identifiable (i.e., no identifiers removed) or de-identified to some degree. The term *anonymized* is commonly used to describe data that have been de-identified very aggressively, to the point that it is impossible for a researcher to link any information back to a specific individual; however, de-identification is not a perfect tool, since there is always some risk, however small, of re-identification. Thus, anonymity cannot be guaranteed. Instead, the only certainty is that any attempt at de-identification places data somewhere along a spectrum involving progressively stronger de-identification, which is correlated with a progressively lower chance of re-identification. The key, therefore, lies in finding an appropriate level of access control for the given level of de-identification. This will require a certain degree of customization when data sharing agreements are prepared to ensure that researchers receive useful data and data custodians are satisfied with the controls that are in place to mitigate re-identification risk (Cavoukian & El Emam, 2011).

Good Practice: Instead of classifying data as identifiable or de-identified, it is useful to view de-identification as a continuum. Because data may be strongly or mildly de-identified, the Panel did not single out one specific process for dealing with de-identified data. Rather, the degree of de-identification should be used to determine the circumstances under which the data may be made accessible for research purposes.

4.3.2 Duties and Roles of Provincial Custodians of Identifiable Health Information

While de-identified data are not subject to privacy laws, identifiable data must be used for many health research objectives (see Chapter 2). For such identifiable health information, provincial laws impose extensive duties on data custodians — to safeguard data in their possession, and, in some provinces, to create information protection practices and data transfer protocols (see Table 4.1, row 3). Consent of research subjects is typically required to collect, use, or disclose personal health information. However, some legislation allows health data collection, use, or disclosure without consent in certain cases, and approved research is one of these cases. Under these laws, researchers must usually show that identifiable information is essential to the research, that the use of data without consent is unlikely to adversely affect the subject individuals, that it is impossible or impracticable to seek individual consent, and that the research is in the public interest. Box 4.2 shows the detailed requirements imposed on custodians in Alberta. Ontario has similar guidelines for information custodians.

Box 4.2
Custodians in Alberta: A Comprehensive Range of Duties

Under the Alberta *Health Information Act* custodians have duties to:

- protect confidentiality of information in their control as well as information disclosed to persons outside Alberta;
- protect against "any reasonably anticipated" threats or hazards to the security of the data or against loss of the information, and against unauthorized access, modification or disclosure of the information;
- implement measures to address "the risks associated with electronic health records" and for the proper disposal of health records;
- identify all of its "affiliates" — employees, people providing services to the custodian, and professionals providing health services — who are also responsible for compliance with the legislation;
- develop policies for facilitating the implementation of the legislation, and provide copies to the Minister; and
- develop privacy impact assessments and submit them to the Minister for approval before making any changes to their administrative practices for handling health information.

These duties applicable to custodians of health information in Alberta are among the strictest in any provincial legislation.

(GOA, 2014b)

Good Practice: To ensure privacy is respected, and to clearly delineate the requirements for access to identifiable data without consent, good practice suggests showing that the research serves the public interest, that obtaining consent is impracticable, that identifiable data are necessary to the research project, and that physical, electronic, software, and all other security measures are appropriately calibrated to protect the data and to sanction any misuses.

4.3.3 Liabilities of Custodians for Breach

To enforce these duties, many provincial laws provide for sanctions if custodians breach their duties under legislation, including those imposed by privacy commissioners and, in some instances, even some form of prosecution (see Table 4.1, row 4). As well, custodians who breach their duties may be subject to tort liability (Govt. of SK, 2005; CAO, 2012; OSCJ, 2014; GNL, 2014; Govt. of BC, 2014a; Govt. of MB, 2014a). However, if data custodians can show that they have made all reasonable efforts to secure the confidentiality of data, they may be protected from liability. No provincial/territorial legislation in Canada defines what "all reasonable efforts" are with any precision. Chapter 5 articulates governance measures that could help define "reasonable efforts."

4.3.4 Provisions to Facilitate Research

As long as custodians have fulfilled their legal duties, and researchers have met conditions for approval of research using identifiable data, all provincial/territorial legislation allows collection, use, or disclosure of identifiable health information without consent for approved research studies (see Table 4.1, row 5). Federally, PIPEDA and Statistics Canada's policies also have similar provisions.[13] However, there is considerable variation among provinces/territories in the specific requirements that researchers must meet to receive or access health information from custodians.

To be eligible, researchers generally must obtain approval from an REB or another responsible entity. The entity responsible for approving proposed research may be the provincial privacy commissioner, a special body with the exclusive power to review research proposals, an REB approved by a provincial oversight body or by the minister of health, or an REB constituted according to specific standards (such as those set out in the TCPS or in privacy statutes themselves), depending on the province (see Table 4.1, row 6). At Statistics Canada, the *de facto* REB is Statistics Canada's Executive Management Board (StatCan, 2014g). Box 4.3 shows the Newfoundland and Labrador approach to research ethics review, in which a single regulatory body has the power to either approve research itself or approve other REBs to carry out this function.

13 CIHI, while a national organization, is governed by Ontario legislation.

Furthermore, in most provinces and at Statistics Canada, researchers must also satisfy a series of criteria to be granted access to data other than strongly de-identified data without consent from participants. In some provinces, the

Box 4.3
Newfoundland and Labrador Health Research Ethics Authority

In 2011, a Health Research Ethics Authority (HREA) was formed to provide central oversight of research ethics review in Newfoundland and Labrador. Legislation in the province provides the same level of protection for health information as that in other provinces, but also has provisions for promoting public trust in health research through the HREA. Researchers may seek approval either from a standing provincial Health Research Ethics Board or from another REB, even out of province, that obtains approval from the HREA. After initial approval of the research proposal, researchers must be overseen by the approving REB and must report back to it when the project is complete (Kosseim *et al.*, 2012; HREA, n.d.).

law is silent on the criteria to be applied for approval, or requires REBs to apply the TCPS criteria. In those provinces that enumerate criteria, these include showing why the researchers need to use identifiable health information, why it is impracticable to obtain individual consent for the information (factors for determining this are found in CIHR's *Best Practices for Protecting Privacy in Health Research* (CIHR, 2005)), and how the benefits of the proposed research outweigh the risks to the confidentiality of the data on the subjects of the study.

In addition, in some provinces, researchers must prepare and submit plans for data security for approval (see Table 4.1, row 7). Researchers must usually sign agreements with custodians requiring them to take steps, extensive in some provinces, to secure the confidentiality of the information they receive (data- or material-transfer agreements) (see Table 4.1, row 8). Box 4.4 shows an example of such researcher-custodian agreements.

Box 4.4
Researcher-Custodian Agreements in Nova Scotia

Nova Scotia's *Personal Health Information Act* (GNS, 2010a) has extensive requirements for the agreements that researchers must sign with data custodians, which are modelled after the requirements in Ontario's *Personal Health Information Protection Act* (PHIPA). The Nova Scotia legislation requires that the researcher:

- comply with any terms and conditions imposed by an REB as well as those imposed by the custodian;
- use the information only for the purposes outlined in the research plan as approved by an REB;
- not publish the information in a form that could be used, either alone or with other information, to identify an individual;
- allow the custodian to access or inspect the researcher's premises to confirm that the researcher is complying with the terms and conditions of the Act and of the agreement between the custodian and the researcher;
- notify the custodian immediately and in writing if the personal health information is stolen, lost, or subject to unauthorized access, use, disclosure, copying, or modification;
- notify the custodian immediately and in writing of any known or suspected breach of the agreement between the custodian and the researcher; and
- not attempt to identify or contact the individuals unless the custodian or researcher has obtained their prior consent.

These requirements are comprehensive and create a wide range of duties for researchers under the resulting agreements.

Good Practice: To ensure that researchers are accountable for protecting data confidentiality, good practice suggests that full and explicit data transfer agreements between researchers and custodians are needed for each research project.

In addition to requirements for transferring data directly to research groups, under some provincial legislation, certain organizations may receive identifiable health information for research purposes (see Table 4.1, row 10). Such "prescribed" or "designated" entities include Cancer Care Ontario, CIHI, ICES, and MCHP. Box 4.5 describes how this arrangement works for one such research centre.

It should be noted that there is a legal grey area concerning whether information is collected or used for research or some other purpose. In most cases, it is clear that information is being collected or disclosed for approved health research. However, much health information is also shared between health facilities, providers, and government agencies to assess or improve quality of care. Whether such disclosure or use constitutes research or should be governed by internal quality assurance mechanisms could become unclear in some instances.

Box 4.5
Disclosure to a Designated Entity: Manitoba Centre for Health Policy at the University of Manitoba

Under the Manitoba legislation, trustees of health information may disclose the information to MCHP or CIHI, both designated as "prescribed health research organizations." Such disclosures must be for the purposes of:

- "analyzing the health status of the population;
- identifying and describing patterns of illness;
- describing and analyzing how health services are used;
- analyzing the availability and adequacy of human resources required to provide health services;
- measuring health system performance; or
- health system planning"

With such broad information-gathering powers, MCHP has built large databases of research data, which researchers may apply to use and which continue to generate studies and analyses of population health and system performance.

(MCHP, 2014e)

4.4 REGULATION OF INTERNATIONAL AND INTERPROVINCIAL DATA SHARING

Sharing health information across borders is subject to legislation and regulation, which vary among countries and provinces. There is some international guidance, such as guidelines issued by the OECD (OECD, 2013c) and the Asia-Pacific Economic Cooperation Cross-border Privacy Enforcement Arrangement (APEC, 2014). However, each country has the primary responsibility for regulating international data flows. The legal barriers in many countries have been cited as the primary obstacle to transnational research studies (Kuipers & van der Hoeven, 2009; Zika *et al.*, 2010; Colledge *et al.*, 2013).

Data sharing involving the United States has raised concerns because of the wide powers of government surveillance under the USA PATRIOT Act (USA PATRIOT Act, 2001) and other laws authorizing surveillance. If health data are shared with U.S.-based recipients or using U.S.-based computer servers, the data could come under the purview of these laws and become subject to interception. This could limit the privacy protections for such data, both in the country where the data are originally entered into the database and in the United States creating a chilling effect on the sharing of health information for research purposes.

Sharing Canadian health information outside of Canada is governed in different ways depending on the jurisdiction (Weisbaum *et al.*, 2005). At the federal level in Canada, Statistics Canada has its own federal legislation, the *Statistics Act*, and a detailed set of policies governing its data sharing agreements (GOC, 2014a). The federal privacy law, PIPEDA, applies primarily to the collection, use, or disclosure of personal information by a private-sector organization in the course of commercial activities. It requires that such data custodians ensure that information recipients in a foreign country provide a level of protection comparable to that found in Canada (GOC, 2014b). As well, the federal Treasury Board has prescribed guidelines for international data sharing (TBS, 2010). At the provincial level, laws in some provinces are silent on international data sharing (GO, 2012; Govt. of MB, 2012, 2014b), while others permit it for approved research purposes (GNS, 2010b; Govt. of BC, 2014c), and still others place additional restrictions on it (GDQ, 2014a, 2014b; GOA, 2014b), including showing that the foreign research entity has legal protections comparable to those in Canada.

For data sharing between provinces within Canada, provincial laws also vary. In several provinces, the legislation is silent on out-of-province disclosures of health information. In others, out-of-province data sharing is permitted if (i) it meets the province's provisions for research, such as approval by an REB or other entity, or (ii) the custodian ensures that the receiving province's laws provide equivalent confidentiality protections for the health information or that the receiving researcher has adequate confidentiality safeguards (see Table 4.1, row 11).

Legal Challenges for Multi-Jurisdictional Data Sharing

The extensive variation in the regulation of cross-border data sharing among countries and among provinces within Canada impedes national or transnational research studies (Cate, 2008). Such studies have great potential for building evidence and knowledge to improve health outcomes, often through the creation of large databanks that pool data from many jurisdictions

(OECD, 2007; Kosseim *et al.*, 2014). Furthermore, technological advances such as fast internet and "cloud-based" data storage, where data are stored on remote servers accessible from any country, have made such studies much more feasible (Seddon & Currie, 2013).

In addition to variation in legislation and regulation, another of the most immediate obstacles to this large-scale research is that research ethics approval in one province or country may not be recognized in another jurisdiction. This could arise from basic differences in legislation or from different interpretations of common terms such as "identifiable," "impracticable," or "all reasonable efforts" to protect data, as discussed earlier in the chapter.

Within Canada specifically, while legislation in some provinces permits approval from REBs or other research ethics approval entities outside the province, in others it does not. Since custodians must comply with the laws of their home province, if that province refuses to acknowledge an approval from an out-of-province REB, the custodian cannot share the health information with out-of-province researchers without further approvals. This can pose a problem, especially when the province where the health data are sought has more extensive oversight of REBs and standards for research approval than the jurisdiction from which the original REB approval came.

Even for studies in which researchers have obtained consent from participants, problems may arise with the use of that consent for interprovincial studies. For example, if researchers located in Alberta wished to use data obtained with consent during an earlier study in another province, the REB in Alberta may not agree to use of data covered by the out-of-province informed consent. Different national and provincial jurisdictions may have different interpretations of the adequacy of informed consent (e.g., in the case of whether a consent given for earlier research is valid for a later study on a related but different subject). The regulator may consider the informed consent inadequate, or the proposed research purpose as too different from the original purpose (Steinsbekk *et al.*, 2013).

Another obstacle to cross-border data sharing is the multitude of agreements and policies with which researchers must comply. Among countries and among provinces within Canada, there can be wide variation in the administrative steps and undertakings that researchers must make with the custodian institutions. In some cases, these agreements or policies are drafted very broadly, to ensure compliance with domestic legal requirements, and can expose researchers to risk (Joly *et al.*, 2011). If any data breaches or other adverse events related

to the shared data occur, researchers would be liable under the agreement for any resulting damage, in addition to being subject to other remedies and penalties under the applicable legislation.

As well, differences in enforcement mechanisms among countries can pose a challenge for the regulation of data access. When data are stored or accessible in multiple jurisdictions, as in cloud-based storage, it may be difficult to know which country's laws apply (Seddon & Currie, 2013). Furthermore, enforcement mechanisms for monitoring the duties of custodians and researchers vary widely. Some countries have few mechanisms, while others have more robust mechanisms. As noted above, Canadian legislation is among the latter, as it commonly provides privacy commissioners with wide powers to investigate and make orders, and creates quasi-criminal offences for breach of the law. Finally, there are also non-legal barriers to cross-border data sharing, including conflicts between the participating institutions' internal policies and procedures.

These barriers to multi-jurisdictional research could be overcome through measures to make regulatory bodies in different countries and provinces more consistent. Common definitions for key terms and recognized templates for data transfer agreements would be useful (Knoppers *et al.*, 2013). It would also be useful to clarify when obtaining consent from individuals is impracticable and when identifiable health information is reasonably needed.

4.5 CANADA'S ETHICAL FRAMEWORK

As well as abiding by the law, researchers must also comply with established ethical standards, which have been articulated in international and national ethics guidelines.

There are two main overlapping sources of research ethics governance in Canada. First, the federal funding agencies impose research ethics review for all federally funded research, guided by their research ethics policy, the TCPS. Second, Health Canada imposes research ethics review through its Clinical Trials Regulations as a pre-condition for allowing clinical trials of drugs or medical devices. Health Canada refers to the International Conference on Harmonisation of Good Clinical Practice Guidelines (ICH-GCP), the World Medical Association's Declaration of Helsinki, and the TCPS for research ethics and clinical trial standards that must be respected in pharmaceutical trials (Hadskis, 2002). In addition, Canadian privacy laws also impose research ethics review as a pre-condition for research involving health data.

However, some research involving human subjects in Canada does not appear to be covered by any of the research ethics requirements (for example, health research conducted in the private sector outside the context of drug or medical device approval). The enforceability of research ethics is also unclear. The federal funding agencies' main mechanism to enforce the TCPS is withdrawal of funding. Health Canada's power to enforce the TCPS, the ICH-GCP, and the Declaration of Helsinki, which are characterized as "guidance" documents, is also limited (Sprumont *et al.*, 2007).

The TCPS has emerged as the main ethical guideline in Canada. Its three core ethical principles — respect for persons, concern for welfare, and justice — govern the access, use, and sharing of research data. Under the first principle, misuse of stored health information may violate individuals' autonomy, which is embedded in the concept of respect for persons. Under the second principle, research can also be seen as a key tool to promote the welfare of individuals and communities. The third principle, justice, requires that research should not unfairly burden some individuals, and that research is aimed at providing potential benefits that will be distributed fairly. The principle of justice also requires eliminating or minimizing any potential stigmatizing impact of health research on specific communities. For example, Aboriginal peoples may be concerned about the potential for such stigmatization from the use of their health data for purposes other than those for which their consent was given (CIHR *et al.*, 2014).

The rest of this section examines four aspects of Canada's ethical framework in light of these ethical principles: ensuring the benefits of research, protecting privacy, consenting to research, and managing risk using proportionality.

4.5.1 Ensuring the Benefits of Research

One of the arguments for enabling research access to data is that resulting research provides important benefits. There are well-established benefits to society from health research and health system innovation, which is one of the basic reasons why governments fund these activities. Supporting scientifically sound, ethical research that will translate into wider social and economic benefits is in the public interest.

Much of the data to advance research and innovation already exist within publicly funded organizations. Taking advantage of these data thus increases activities in the public interest. The Office of the Australian Information Commissioner, for example, has made clear that public-sector information is a national resource that should, wherever possible and appropriate, be made available for community access and use (Adams & Allen, 2014). Similarly, in

Canada, the Information and Privacy Commissioner of Ontario has promoted the use of health information for wider public interest purposes beyond immediate patient care and argued for enhanced data protection measures to facilitate and regulate such uses (Cavoukian & Alvarez, 2012).

In fact, it is increasingly argued that there is an ethical duty to contribute to research. Reducing the risk of harm for individual patients, for example, by providing evidence of the adverse impact of a medical treatment or practice, is an ethically defensible course of action.[14] However, in addition to simply avoiding harm, some researchers are calling for proactive measures to maximize well-being, and suggest that it may be unethical not to make use of existing data sources to help individuals, communities, and future generations who could readily benefit (Stanley & Meslin, 2007). When information in existing data holdings could be used to reduce premature death, or to improve survival of a disease, failure to use those data would create, or at a minimum fail to address, specific harms (Allen *et al.*, 2013).

Some have argued that those who benefit from publicly funded health care should also contribute to it to protect others (Meslin & Cho, 2010; Forsberg *et al.*, 2014). For example, when new therapies become available on the market, there is limited evidence from the regulatory approval process concerning their efficacy and safety; reports of adverse effects may accumulate as the therapy is used. Society at large would benefit from being able to intervene in the cases where accumulating post-marketing evidence suggests that there is a safety concern (Sethi, 2014). Hence, there is an ethical imperative in such circumstances to intervene, and to do so as soon as possible. Data reporting, sharing, and linkage allow this critical information to come to light.

In such cases, failure to analyze data and prevent future harm could be unethical. Indeed, those who have suffered harm from a failed drug, for instance, may not want their suffering to be in vain, and therefore expect lessons to be learned. This is particularly true when patients have consented to the use of their data for research purposes. In the context of a rare disease, for example, it may be difficult for researchers to gather sufficient data to do meaningful studies that may assist in fighting the disease. In such circumstances, the ethical imperative is to make as much use as possible of information, including through data linkage, "to maximize the chance that patients' contributions translate into therapeutic advances" (Kush & Goldman, 2014). This aligns with the ethical principle of respect for persons.

14 Non-maleficence, or the duty to avoid, prevent or minimize harms to others, is captured in the concept of welfare in the TCPS, along with beneficence (doing good) and proportionality (CIHR *et al.*, 2005).

A hospital in England provides a real-world example of the ethical obligation to examine data. High mortality rates at the Stafford Hospital, eventually uncovered through data analysis, suggested that many hundreds more individuals had died than would have been expected between 2005 and 2008 as a result of poor medical and management practices (Holmes, 2013). Although significant and direct physical harm had been done by the management and practitioners in the hospital, failure to analyze the data promptly had also caused preventable harm. This is at odds with the ethical principle of concern for welfare. Among the conclusions of one of the ensuing public inquiries was the following:

> All such organisations have the responsibility to detect and redress deficiencies in local management and performance where these occur. […] not just the [hospital]'s Board but the system as a whole failed in its most essential duty – to protect patients from unacceptable risks of harm and from unacceptable, and in some cases inhumane, treatment that should never be tolerated in any hospital.
>
> (Francis, 2013b)

Furthermore,

> All professionals, individually and collectively, should be obliged to take part in the development, use and publication of more sophisticated measurements of the effectiveness of what they do, and of their compliance with fundamental standards. Patients, the public, employers, commissioners and regulators need access to accurate, comparable and timely information.
>
> (Francis, 2013a)

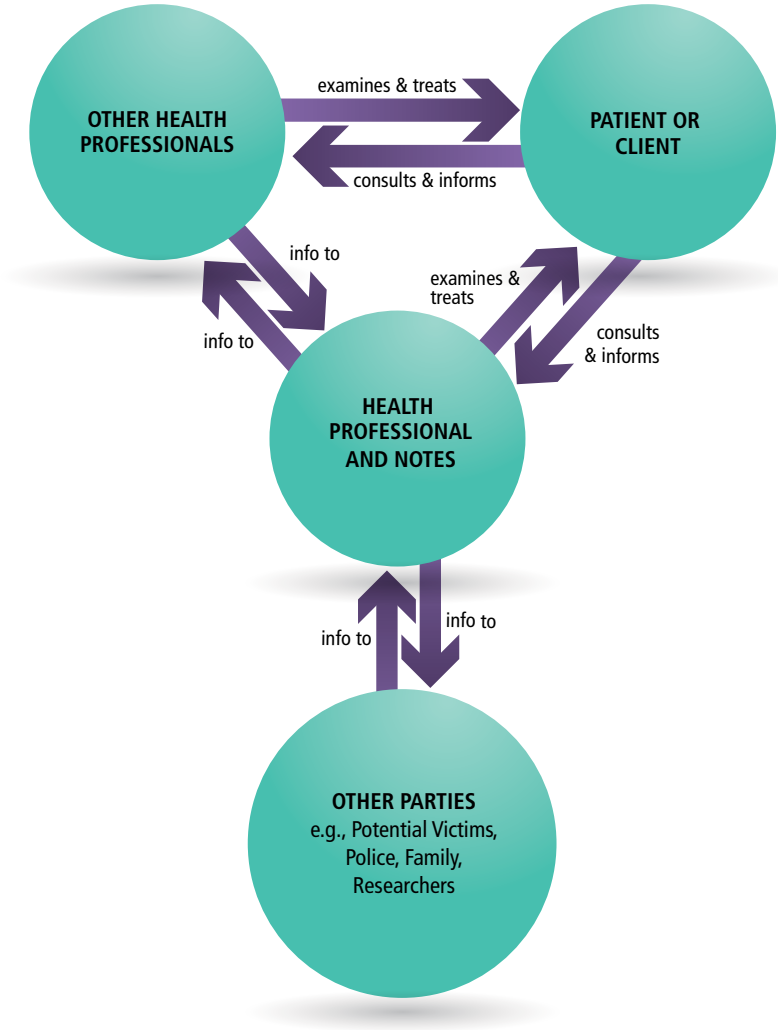
The impacts of new drugs and practices on all parts of society also need to be considered. Capturing all impacts may be feasible only through analysis across databases. An example mentioned above is ADRs, which can often be detected only from large amounts of data, through data pooling, conducted effectively and in a timely fashion.

4.5.2 Protecting Privacy

As noted in Section 1.2, the Panel distinguishes between the concepts of privacy and confidentiality. *Privacy* broadly encompasses protection of one's physical self, protection of one's private physical space, and protection of information about oneself and one's activities (SCC, 2004). *Confidentiality* refers to the duties and practices of people and organizations to ensure that individuals' personal information only flows from one entity to another according to legislated or otherwise broadly accepted norms and policies. Respect for privacy has emerged as an ethical standard for research. The TCPS highlights harms to individuals that may result from inappropriate use and disclosure of personal information, such as discrimination in insurance and employment (CIHR *et al.*, 2014).

Protection of confidentiality is relevant to the public during interactions with health and other professionals. For example, when patients communicate with their doctor, order a drug prescription from their pharmacy, or enter a clinic that specializes in a specific disease, they are revealing personal information with an expectation that it will be used within that context and kept confidential. In these situations, professionals such as doctors and pharmacists have specific legal obligations to keep that information confidential and not to reveal it to others without consent of the individual or as specified in law. However, obligations of confidentiality are never absolute, and laws and ethical guidelines provide exceptions to this duty of confidentiality, allowing the sharing of information in a variety of specific non-research circumstances without informed consent such as public health emergencies or exceptional police powers. Figure 4.1 illustrates the various relationships that can give rise to an expectation of confidentiality.

What constitutes a confidentiality violation in the context of accessing individual data for research purposes? Confidentiality is clearly breached when third parties surreptitiously hack into databases and use that information, for example, to blackmail or deceive individuals. But mere access to information without consent, many would argue, does not necessarily violate confidentiality. Health information stored in databases is accessed by many people without explicit consent for various legitimate purposes, and this is acknowledged in privacy laws. Moreover, privacy law also relies on a concept of "reasonable expectation of privacy." A person's privacy is therefore violated in law only if it was reasonable for the person to expect that information would not be shared or accessible.



Adapted with permission: Dawson (2006)

Figure 4.1

Common Confidentiality Relationships in Health Care

In relationships in health care, an individual (patient) has interactions with a health professional, who takes notes and enters data on the encounter. The primary contact with a health professional may lead to additional encounters with other health professionals (pharmacist, specialist) with whom confidential information may be shared. In exceptional circumstances, usually mandated by law, the health professional may share information with others outside these confidential relationships. For example, most jurisdictions have laws that include duties to warn a potential victim, to inform police of criminal acts, to inform family when the patient is not competent or a proxy decision-maker is needed, and to provide data for research.

4.5.3 Consenting to Research

The early history of research involved various episodes with catastrophic consequences to patients resulting from a lack of respect for consent to medical research. These episodes included the syphilis experiments in the United States from 1932 to 1972 (Rusert, 2009), and atrocities of World War II, including forced human experimentation (Grodin & Annas, 1996). Public attitude surveys consistently find that the public wants to be asked and to have a right to make choices on the use of their bodies (Dixon-Woods & Tarrant, 2009; Murphy *et al.*, 2009; Lewis *et al.*, 2013). Consent regarding use of personal information has evolved more recently, and carries no intrinsic risk of physical harm.

In the context of consent regarding the use of a person's data, there are different kinds of consent, ranging from specific to broad. When individuals give a *specific* consent for research using their data, only the data described in the consent can be shared and only under the conditions accepted, such as, for example, research on a particular disease or the use of data in a certain structure or certain circumstances. By contrast, under a *broad* consent, individuals may agree to research on a wide range of health conditions or even unspecified future biomedical research (subject to ethics or other kinds of approval) (Kosseim & Jospe, 2011). Again, the conditions of data sharing and the data confidentiality and security methods may be described in the consent and thus apply to the future use of the data.

Some suggest that the right to control future use of health information is part of privacy protection, while others believe it is part of personal autonomy (Pritts, 2008). Regardless of the nature of the interest, approaches to this right to control vary from more restrictive to more flexible. A very restrictive view is that people should provide consent for every specific future use of data for which they did not originally give their explicit informed consent. A more flexible view is that autonomy is respected when data are used for research-related purposes in line with the type of research for which people originally provided consent.

The restrictive view is increasingly questioned, particularly in the context of biobanks, where participants have specifically consented to provide samples and related health information for ethics-approved research purposes that are not yet known at the time of storage (Knoppers *et al.*, 2007). Moreover, there is strong evidence that requiring consent before making data available for research is likely to cause serious bias in the data, for example if those who withhold consent are systematically different in a health or biologically relevant respect from those who do provide consent. As a result, in an important range of circumstances, essentially where a government agency with appropriate

legislative safeguards is the data custodian, "consent" is deemed to have been given. For example, in the health surveys conducted by Statistics Canada since 1994, respondents (close to one million) are asked to consent to linking their survey responses to provincial health-care records, and to share their data with provincial health ministries; well over 90% do provide their consent (StatCan, 2004). However, they are not asked for their consent to share their data with researchers. Rather, researchers' access to the data is governed by the *Statistics Act* and the policies of Statistics Canada, as described in Section 2.5.3.

Informed consent also seems an inadequate mechanism to deal with various other components of research involving health information. The concerns raised by commercialization of research findings and access to the benefits of research to research subjects, patients, families, and communities are hard to deal with through individual consent. The familial and communal components of much information, particularly genetic information, make it clear that research has implications for others who are not directly involved in individual informed consent procedures. Concerns about stigmatization or discrimination of communities (e.g., Aboriginal groups) cannot adequately be dealt with through individual consent.

For specific diseases, new initiatives are being developed to more actively involve patients and research subjects in the use of their health data or related biological samples for such research. These initiatives aim at furthering autonomy. For example, "dynamic consent" refers to an ongoing consent process in which individuals remain involved in the decision-making process for future use of samples (e.g., Kaye *et al.* (2014)). Similarly, data sharing itself contributes to autonomy in the ways patients can make decisions about their information. Particularly in the context of ADRs to pharmaceuticals, adequate data sharing is essential for reliable drug safety and effectiveness research. Patients' informed decision-making (and thus their autonomy) is enhanced when they have access to reliable information, for example, on when a given drug is likely to be helpful or not. In Europe, the European Court of Human Rights has stated in several instances that access to relevant health information can be seen as a component of the fundamental right to the protection of private and family life (Lemmens, 2013).

There may be no confidentiality risk associated with the sharing of data (for example, in the case of properly and adequately de-identified data), making it unnecessary — from a privacy perspective — to limit data access. It may be very difficult to recontact individuals to obtain consent for their data to be used for research if that consent was not obtained when the data were initially gathered. And yet, these individuals may be willing to have their data used

for research. Furthermore, if a sizeable proportion of the individuals do not give consent, the quality of the data for analysis will be reduced, possibly to the point of uselessness (Tu *et al.*, 2004). The results may even be misleading, and thus may lead to harms for others (Tu *et al.*, 2004; El Emam *et al.*, 2007; Kho *et al.*, 2009; El Emam, 2013a).

Some commentators further argue that consent cannot fulfil its purpose of ensuring that research participants are fully informed and agree to use of their data. Studies have shown that people sign consent forms without really knowing what they are signing, and so consent forms may not offer an ethical defence for data custodians. McDonald and Cranor (2008) estimated that the time spent reading privacy notices each time an average person visited a new website would amount to 244 hours per year (at an average of 10 minutes per policy). Given this time burden, the expression of consent to such notices may not mean the user has read the notice and made an informed choice. The ethics of, and trust in, those holding confidential data may be called into question by the public, regardless of whether they have gained consent, they have released data, or they are allowed to hold data under legislation.[15]

The process of obtaining consent by a clinician or researcher can be useful to raise awareness of the use of data and the benefits that research could entail, even if consent is not required. As well, if obtaining consent is not needed or practicable, alternatives to inform individuals, such as communication and notification strategies, should be considered when using health data.

Given the challenges with regard to determining the appropriate role for consent, including the risks of bias and the extent to which in practice it can ever be fully informed, this is an area where, in a democracy, the balance is generally struck in legislation and regulation. The result almost universally is that while consent is often required, just as often there are situations enumerated where it is not required. From a broader perspective, therefore, the challenge is how best to enumerate and define those situations where consent to use individuals' identifiable or potentially identifiable information is not required.

15 Health research that improves public health generates a benefit for all (Anomaly, 2011). Public goods, such as the safety of medicines, cannot be tailored to individual demand. Consequently, it is difficult to justify a specific consent-based approach to their provision (O'Neill, 2003, 2004). As a result, there is no ethically defensible right to consent or to withhold consent in the same way that there are moral and legal rights to privacy and confidentiality (Clark & Weale, 2011). For many public goods, sufficient provision is enabled through notification with an opportunity to opt-out.

To address the limitations of informed consent and additional issues at stake in sharing information, privacy governance is needed to determine what constitutes appropriate and ethical use of data (see examples in Chapter 5). All Canadian privacy laws require governance as an essential tool to promote and protect privacy and other social goals. Indeed, in the context of the development of databases and biobanks, participating individuals may "consent to a governance system" when their information is collected and stored, allowing future use of data determined through the governance system (Austin & Lemmens, 2009).

In regard to governance requirements, Canadian laws were modelled in part on Fair Information Practices (FIPs), a governance framework developed by the OECD beginning in 1980 to promote privacy protection. FIPs were a response to concern about the risk of privacy loss when digital data started to become available in all walks of life in the 1970s. Since that time, FIPs have been translated into the legal systems of many countries, including Canada (Cavoukian & Chanliau, 2013). In 2013, the OECD refined and expanded its guidelines on privacy protection, which include the FIPs, to take into account the proliferation of available data and the increased potential benefits to research (OECD, 2013b).

A key element of the new guidelines is recognition that sole reliance on consent for data use may be impracticable and insufficient to protect individuals' privacy. Furthermore, limiting data collection to the specific purpose in the notice used to solicit consent inhibits their future use for the benefit of society at large. The OECD notes:

> The notice and consent system, on which data collectors and data users have come to rely, was designed to empower individuals to make decisions about their personal data, but the evolution of data collection and data use has severely weakened that power while imposing increasing burdens on data subjects and on society. While notice and consent may provide meaningful privacy protection in appropriate contexts, this approach is increasingly ineffective as the primary mechanism for ensuring data privacy.
>
> (OECD, 2013a)

As a result, the revised approach "shift[s] responsibility [for information privacy protection] away from individuals and toward data collectors and data users, who should be held accountable for how they manage data rather than whether they obtain individual consent" (OECD, 2013a).

Key elements of the new guidelines include:

- focusing on practical assessment of the benefits and risks associated with data uses;
- restoring balance between privacy and the free flow of information, and avoiding suppression of innovation with overly restrictive or inflexible data privacy laws; and
- putting in place practical frameworks and processes for identifying, balancing, and mitigating harms from inappropriate uses of personal data.

(OECD, 2013a)

International evidence suggests that data useful to research should be accessible, but that there should be a greater onus on those who collect and use data to protect data confidentiality, which in turn calls for stronger governance practices, as discussed in Chapter 5.

Governance principles can also address some gaps in the legal and ethical frameworks. In some cases, long-standing principles of confidentiality in common law readily incorporate protecting data confidentiality. However, to allow for future technological or social change, the law is not always specific. In addition, principles of informed consent, which are focused on individual interests, may be limited in their ability to deal with collective social, political, and cultural interests of specific discrete communities such as Aboriginal groups. For these reasons, governance mechanisms that account for such realities in the context of privacy protection are vital.

4.5.4 Managing Risk Through Proportionality

Proportionality — keeping measures to protect privacy proportional to the risks of harm from uses of individual information — is a central element in the TCPS. Furthermore, considerations of proportionality permeate legal rules about privacy protection: privacy is not an absolute right in the sense that it needs to be protected at all costs. Interference with privacy can be justifiable so long as the degree of interference is proportionate to the wider social benefit and is needed to achieve that social benefit. In practice, then, use of individual data for research, even without consent, can be justifiable. In some cases (e.g., for a university-based researcher), the onus is on those who seek to use the data to justify this use. This is particularly true if consent will not be sought. Moreover, a proportionate approach ensures that appropriate scrutiny is tailored to the level of risk presented by the research. Thus, the ethical acceptability of the research should involve consideration of the foreseeable risks, the potential

benefits, and the ethical implications of the research (CIHR *et al.*, 2014). In other cases, (e.g., for data held by Statistics Canada), legislation is already in place to deem consent unnecessary.

Proportionality applies to the real risks inherent in data sharing, which can include risks to reputation of undertaking data linkage and sharing, risks to privacy (which cannot be completely eliminated),[16] and other risks specific to particular types of data. El Emam and Arbuckle (2013) point out that de-identification of genetic data would render the data useless for research purposes because the data would have to be severely distorted to become non-identifiable (see also Austin and Lemmens (2009)). Therefore, research using genetic data involves a particular risk of identification of people with genetic diseases.

As a result, appropriate practice suggests that in those situations where REB review, rather than legislation, is the primary method for judging these important trade-offs between individual privacy and the public good benefits of the research, ethics review must include this kind of risk assessment. Without such an assessment, any policies or practices required to address those risks may be too little or too great, and, hence, disproportionate. Analytical judgment should include appropriate consideration of risks against a background of core objectives so that an appropriate weight can be applied (Sethi & Laurie, 2013).

Good Practice: The TCPS recognizes that risk cannot be eliminated but should be considered proportionately. Good practice suggests incorporating risk management in all aspects of governance, including ethical governance.

4.6 CONCLUSION

Protecting privacy and confidentiality is a legal and ethical duty. Yet, at the same time, there is an equally important ethical duty to conduct research through access to data to create public benefits and avoid future harm. With effective mechanisms to protect privacy and confidentiality — such as a carefully calibrated combination of a particular degree of de-identification and a corresponding level of access security — an ethical framework such as the TCPS clearly supports enabling access to data for research purposes.

16 The U.K. Academy of Medical Sciences has twice reported on the need to strike a better balance between privacy and proportionate governance (2006 and 2011) (AMS, 2011).

The legal frameworks in Canada demonstrate a clear and consistent overall intent to balance the public interest in research with the protection of privacy and confidentiality. There are, for instance, no legal obstacles to releasing data that are considered de-identified under a given province's legislation to researchers. There are many examples in Canadian legislation where research and analysis using identifiable individual data are specifically permitted, and there are agencies that publicly fund this research.

However, parts of Canada's ethical and legal framework lack clarity. There are variations in approach and in definitions of key terms in the laws across provinces and the federal jurisdiction. Consequently, the development of data sharing guidelines based on the terms defined in these laws can be confusing. Further, with the rapid technological changes in the capacity to move large volumes of data, as well as the explosion in the volume of available data, the various legislative and other approaches are often seriously lagging. The result is significant uncertainty as to what specific practices are allowed in various circumstances. To deal with this uncertainty, it is useful to view data on a spectrum, beginning with fully identifiable data and moving towards data that are more strongly de-identified and less likely to be re-identified. The key lies in finding an appropriate level of access control for the given level of de-identification.

Table 4.1

Provincial Legislative Provisions Governing Health Information Privacy and Research Promotion

Row		British Columbia	Alberta	Saskatchewan
1	De-identified data can be used freely	Yes	Yes	Yes
2	Definition of "identifiable" health information	Used but not defined	If identity of person is "readily ascertainable"	Identification is reasonably foreseeable from combination of data
3	Custodian duties to safeguard data	General duty to take steps to ensure confidentiality	Extensive duties to develop and follow information security protocols	Must develop information security practices
4	Custodian liabilities for data breaches	Investigation by Privacy Commissioner, possible orders by: • Tort liability, including statutory invasion of privacy tort • Possible criminal prosecution	Investigation by Privacy Commissioner, possible orders by: • Tort liability • Possible criminal prosecution	Investigation by Privacy Commissioner, recommendations by: • Tort liability, including statutory invasion of privacy tort • Possible criminal prosecution
5	Data may be used for approved research purposes	Yes	Yes	Yes
6	Approving entity	FIPPA: Privacy Commissioner must approve; PIPA: No entity designated	Designated REBs under Regulation	REB approved by Minister
7	Criteria for approval decisions	Brief and generally stated	Lengthy, detailed and/or elaborate legislative standards	Lengthy, detailed and/or elaborate legislative standards
8	Researcher-custodian agreements required	Yes, general duty to get agreements	Yes, with extensive and detailed terms	Yes, with extensive and detailed terms
9	Duties of researchers	Researchers not bound by same duties as custodians	Researchers not "custodians"	Researchers not "trustees", but recipients of health information take on all duties of custodian that disclosed it
10	Designated research entities	None	"health information repository"	None
11	Disclosures to another province for research	Permitted if for approved research	Permitted if custodian enters into agreement with researcher(s) that binds the researchers to protect the confidentiality of the data	No restrictions

continued on next page

Row		Manitoba	Ontario	Quebec
1	De-identified data can be used freely	Yes	Yes	Yes
2	Definition of "identifiable" health information	If the data "allows" identification	Identification is reasonably foreseeable from combination of data	If the data "allows" identification
3	Custodian duties to safeguard data	Must develop information security practices	Must develop information security practices	General duty to take steps to ensure confidentiality
4	Custodian liabilities for data breaches	Investigation by Privacy Commissioner, recommendations by: • Tort liability, including statutory invasion of privacy tort • Possible criminal prosecution	Investigation by Privacy Commissioner, possible orders by: • Tort liability • Possible criminal prosecution	Investigation by Privacy Commissioner, possible orders by: • Tort liability, including statutory invasion of privacy tort • Possible criminal prosecution
5	Data may be used for approved research purposes	Yes	Yes	Yes
6	Approving entity	Designated entity, or other REB that meets statutory test	REB not needing pre- approval but meeting statutory test	Privacy Commissioner
7	Criteria for approval decisions	Lengthy, detailed and/or elaborate legislative standards	Lengthy, detailed and/or elaborate legislative standards	Brief and generally stated
8	Researcher-custodian agreements required	Yes, with extensive and detailed terms	Yes, general duty to get agreements	None required
9	Duties of researchers	Researchers not "trustees," unless a prescribed Health Research Organization	Researchers not "custodians"	Researchers not bound by same duties as custodians
10	Designated research entities	MCHP, CIHI	ICES, CIHI, CCO, POGO	None
11	Disclosures to another province for research	No restrictions	No restrictions	Permitted if receiving jurisdiction has equivalent privacy protections

continued on next page

Row		New Brunswick	Nova Scotia	Prince Edward Island	Newfoundland & Labrador
1	De-identified data can be used freely	Yes	Yes	Yes	Yes
2	Definition of "identifiable" health information	Identification is reasonably foreseeable from combination of data	Identification is reasonably foreseeable from combination of data	Used but not defined	Identification is reasonably foreseeable from combination of data
3	Custodian duties to safeguard data	Must develop information security practices	Must develop information security practices	General duty to take steps to ensure confidentiality	General duty to take steps to ensure confidentiality
4	Custodian liabilities for data breaches	Investigation by Privacy Commissioner, recommendations by: • Tort liability • Possible criminal prosecution	Investigation by Privacy Commissioner, recommendations by: • Tort liability • Possible criminal prosecution	Investigation by Privacy Commissioner, possible orders by: • Tort liability • Possible criminal prosecution	Investigation by Privacy Commissioner, possible orders by: • Tort liability, including statutory invasion of privacy tort • Possible criminal prosecution
5	Data may be used for approved research purposes	Yes	Yes	Yes	Yes
6	Approving entity	REB not needing pre- approval but meeting statutory test	REB not needing pre- approval but meeting statutory test	Privacy Commissioner	Designated entity or REB approved by provincial HREA
7	Criteria for approval decisions	Lengthy, detailed and/or elaborate legislative standards	Lengthy, detailed and/or elaborate legislative standards	Brief and generally stated	Brief and generally stated
8	Researcher-custodian agreements required	Yes, with extensive and detailed terms	Yes, with extensive and detailed terms	Yes, general duty to get agreements	None required
9	Duties of researchers	"Custodian" includes researchers	Researchers not "custodians"	Researchers not bound by same duties as custodians	Researchers not "custodians"
10	Designated research entities	None	None	None	None
11	Disclosures to another province for research	Permitted if for approved research	Permitted if for approved research	No restrictions	Permitted if for approved research

5

Effective Governance for Accessing Health and Health-Related Data

- Principles of Good Governance and Best Practice Entities
- Cross-Cutting Aspects of Governance
- Allocating Responsibilities for Aspects of Governance
- Conclusion

5 Effective Governance for Accessing Health and Health-Related Data

Key Findings

- Many institutions, organizations, programs, and activities across Canada are collectively responsible for the provision of timely access to health and health-related data. However, they are only loosely coordinated, even within a single province, let alone across provinces or legal jurisdictions. They are best thought of as a "complex environment of heterogeneous entities" — a collection of diverse, multiple, interconnected entities and elements.
- Not all entities within this complex environment in Canada are formally constituted as discrete organizations with publically accountable boards of directors or their equivalent. Nonetheless, each program, department, project, individual, or organization has a responsibility to follow principles of good governance.
- The Panel selected six entities that succeed in enabling timely access to data for researchers. The governance approaches of these six "best practice entities" share four common principles: enabling appropriate use of data, managing risk, respecting privacy, and maintaining public trust by providing evidence of trustworthiness.
- These best practice entities, within complex environments in Canada or in other countries with similar social and legal systems, collectively address four cross-cutting aspects of governance: privacy, research, information, and network governance.
- A final overarching consideration in delivering good governance is the need to consider the proportionality of governance mechanisms working together, that is, that the level of scrutiny and control over access must be commensurate with the level of risks at stake. Lower risks justify lighter touch governance.

Timely access to data depends on the effectiveness of numerous organizations and other entities dealing with health and health-related data, and on the approach to governance implemented by them. Although often described as a "system," these institutions, organizations, programs, and activities are only loosely coordinated, even within a given province. Yet, they share responsibility for the provision of timely access to health and health-related data. They are better described as constituting a "complex environment:" a collection of diverse, multiple, interacting entities and elements that have the capacity to adapt and learn from experience, but can equally become uncoordinated and inefficient.

In the context of timely access to health and health-related data, the role of governance is crucial in bringing some order to the underlying complexity. For the purposes of this assessment, governance can be thought of as the role of an organization's board of directors (or equivalent) that defines that organization's purpose and develops the strategies, objectives, values, and policies to pursue that purpose. It includes such management tools as mission statements, statements of organizational objectives and values, logic models, organizational performance metrics, risk management frameworks, policies and guidelines for financial and operational matters, stakeholder relations, and the like. As will be discussed later, lying on top of such organizational governance matters are the higher-level, cross-cutting governance issues relating to how the environment as a whole addresses important issues, such as privacy, through the activities of the separate organizations.

The Panel's analysis suggests that those responsible for governance of the organizations dealing with health and health-related data can learn from other organizations within Canada and around the world that have implemented effective governance within their own contexts. This learning would help each component of the complex environment to improve its governance, and thus improve the provision of timely access to health and health-related data.

5.1 PRINCIPLES OF GOOD GOVERNANCE AND BEST PRACTICE ENTITIES

The legal and ethical framework defined in Chapter 4 provides general rules that constrain those making decisions on accessing and providing access to health and health-related data. Transforming those rules into concrete and effective guidance requires modes of governance that address the various layers of complexity within the system. The best such systems are built upon principles.

Principles of good governance in areas of public service have received increasing attention in recent years, both in Canada and elsewhere.[17] One of the most compelling statements of such principles was developed by the Langlands Commission in the United Kingdom (Box 5.1).

17 For example, the Government of Canada's Office of the Superintendent of Financial Institutions was established to contribute to public confidence in the Canadian financial system. It is guided by seven key principles and a supervisory framework (OSFI, 2010).

Box 5.1
Principles of Good Governance

According to the Langlands Commission, several principles are needed for good governance in public service:

1. Focusing on the organization's purpose and on outcomes for citizens and service users;
2. Performing effectively in clearly defined functions and roles;
3. Promoting values for the whole organization and demonstrating the values of good governance through behaviour;
4. Taking informed, transparent decisions and managing risk;
5. Developing the capacity and capability of the governing body to be effective; and
6. Engaging stakeholders and making accountability real.

(OPM & CIPFA, 2004)

Part of the value of this statement is that it can be applied across the entire public sector. Applying these principles to the Canadian context, for example, would suggest that REBs that review data-intensive projects without members who have expertise in the potential impact on privacy would be counter to principles 4 and 5. Overlapping or multiple reviews would be inconsistent with principle 2. Lack of transparency in explaining decisions on preventing access to data would fail to meet principle 6. In the Panel's experience, many of these principles, and especially principle 6, remain goals to aspire to in Canada in the context of timely access to health-related data for research.

The Panel looked for evidence from Canada and internationally of organizations that support timely access to data, and follow good governance practices according to the Langlands criteria. Six institutions and programs (entities) were selected (three from Canada and three from other countries). All have successfully enabled timely access to health and health-related data as well as data linkage. The Panel regards these as "best practice entities" (Box 5.2).

Box 5.2
Best Practice Entities

Farr Institute @ Scotland: The Farr Institute @ Scotland builds on the expertise, infrastructure and established cross-sectoral collaboration developed by the Scottish Informatics Programme (SHIP). SHIP was a research platform for the collation, management, dissemination, and analysis of electronic patient records from across Scotland. It was operated by a network of universities in collaboration with the National Health Service (NHS) Scotland, and funded by the main U.K. granting agencies. It included a new research institute specifically for research based on EHRs and an online research portal providing rapid, secure access to health data. It also found mechanisms to link these data to large third-party data sets. The programme has ended, and its platform has moved to the Farr Institute @ Scotland (SHIP, n.d.-a, n.d.-b).

Wales Secure Anonymised Information Linkage Databank (SAIL): SAIL is a database of anonymized health and social data about the population across Wales, funded by the Welsh Government's National Institute of Social Care and Health Research. The individual-level data sets are held at Swansea University and can be linked together for research purposes, subject to privacy legislation and research approvals (SAIL, 2014).

Data Linkage Western Australia (Data Linkage WA): This data linkage system was established in 1995 to connect all health and related information for the Western Australian population. The system is a coordinated and collaborative effort between the Department of Health Western Australia, two universities, and an institute for child health research. The information is used for research, subject to ethical approval, and for planning to improve the health and well-being of citizens in the region (Data Linkage WA, 2014b).

Ontario – Institute for Clinical Evaluative Sciences (ICES): ICES is a research institute housing a secure yet accessible array of health information on Ontarians. It also involves a community of researchers conducting health services research by linking data from many sources to obtain a comprehensive view of health care and health-care delivery. The majority of data come from the publicly funded health-care system and include medical records, laboratory results, and medical imaging. A "prescribed entity" that may receive individual-level identifiable information under Ontario legislation, ICES de-identifies and anonymizes data collected, restricts access to data to relevant projects, and trains all of its scientists in privacy policy and practice (ICES, 2014a, 2014e).

continued on next page

Ontario – Better Outcomes Registry and Network (BORN): The registry was started in 2009 to collect, interpret, share, and protect data about pregnancy, birth, and childhood in the province of Ontario. Its information system collects data on every birth and young child in the province from hospitals, laboratories, midwifery practice groups, and clinical programs. Its objectives include linking information and providers to address gaps in care, and to support research and innovation to improve maternal and child health. The organization is funded by the Ontario Ministry of Health and Long-Term Care and administered by the Children's Hospital of Eastern Ontario (BORN, 2014).

Manitoba Centre for Health Policy (MCHP): Located in the Faculty of Medicine at the University of Manitoba, the centre is essentially a research unit focusing on population health. Most of its research is from its data repository, containing anonymized individual-level data on Manitobans' use of hospital care, physician services, home care, nursing homes, prescriptions, education, and family services. It also links these data to non-personal (aggregated) socio-economic data from Statistics Canada. The provincial Ministry of Health has funded several large research projects at the centre, accounting for approximately half of its total funding. The rest comes from projects approved by other funding bodies and granting agencies (MCHP, 2014b).

The Panel identified four principles common to all of the entities identified in Box 5.2:

i. **Enabling appropriate use of data** to enhance public well-being, which is one of the key elements of good data stewardship;
ii. **Managing risk** by identifying the range of risks involved in providing data access and minimizing those risks where possible, while acknowledging that risks cannot be entirely eliminated;
iii. **Respecting privacy** to reassure citizens that risks to their core personal interests are kept to an absolute minimum, which is another key element of good data stewardship; and
iv. **Maintaining public trust by providing evidence of trustworthiness**, recognizing that the public has an interest in seeing the confidentiality of their personal data being maintained, but also in seeing that appropriate use is made of health and health-related data for research with demonstrable social value. (This applies to both the data custodians and the researchers involved.)

These principles guide all aspects of governance that these entities have implemented, although specific policies and practices may differ from one institution or context to another. The value of a principle-based approach is that principles are starting points for deliberation and action; they provide a common framework to achieve certain goals while allowing flexibility in the specific means to achieve those goals. Each entity must choose its own approach; however, explicit commitment to these principles helps to ensure that each approach is the best one for each entity in the pursuit of providing timely and effective access to data.

It is notable that all of these six best practice entities are at Level 1; they all operate within a single jurisdiction. At Level 2, best practices are most clearly evident at Statistics Canada and CIHI. At Level 3, the Panel was challenged to find best practice entities. This is a context where practices are evolving — most rapidly in genetic research.

5.1.1 Enabling Appropriate Use of Data

When an entity enables appropriate use of data, it ensures that "all relevant [g]overnance requirements are met, but that [g]overnance is not unduly burdensome to research" (Farr Institute, 2015). As discussed earlier, the appropriate use of health and health-related data enhances social well-being. As a result, in many cases there is an ethical imperative to promote the use of such data. The Panel found that data custodians that had accorded a high priority to the use of data were organized and funded to facilitate access and had developed processes to foster such access.

Clarifying the role of data custodians in promoting the use of data enables appropriate decisions to be made. In some cases, data custodians originally established as secure repositories of data may not perceive supporting access to data for research as part of their core mandate. In most cases, provincial law designates who is a data custodian and prescribes rules for research sharing. However, some data custodians *do* state clearly in their mission statements that enabling appropriate use of data is a core purpose of the organization. This is also in keeping with the "setting the context" principle for risk management (Box 5.3), and the conclusions of the *Second Caldicott Report* in England and Wales (Box 5.4). The mission statement of ICES, for example, states that extensive use of data is intended: its mission is "research excellence resulting in trusted evidence that makes policy better, health-care stronger and people healthier," and its vision is to be "a world-leading institute where data and discovery

improve health and health care" (ICES, 2014a). To complement statements such as these, data custodians could publish regularly how much access they have provided to researchers along with a list of subsequent publications (MCHP, 2010). As well, they could publish details of the researchers who have sought access, a summary of the proposed research, and an account of the intended public interest that the research is designed to address. Statistics Canada, for example, provides information on all approved record linkages on its website routinely (StatCan, 2014a). Population Data BC maintains a publicly available database, which lists all research projects and includes the data sets that were requested for each one (PopData BC, 2015a). Such transparency about data use also fulfils the principle of maintaining public trust.

Best Practice: The appropriate provision of data to researchers is central to those entities highlighted by the Panel. For example, entities such as the Farr Institute @ Scotland, SAIL, Data Linkage WA, ICES, and MCHP clearly state in their mission statements that enabling appropriate use of data for research is a core purpose of their organization.

For some of these best practice entities, strong leadership has enabled them to move from a "data custodianship model," in which holding and securing data are emphasized to the exclusion of other considerations, to a "data stewardship" model, in which enabling access is a core institutional objective. The advisory board for MCHP, for example, includes seven deputy ministers from the Government of Manitoba (MCHP, 2014a). In Western Australia, the Director General (equivalent to deputy minister in Canada) of the Western Australia Health Department (WA Health) is the delegated owner of all data stored, used, and disclosed. As the Director General sets the overall strategic direction of the WA Health, the data custodians "have delegated responsibility for setting the overall strategic direction of the specific data collection to ensure the collection is developed, maintained and utilized in accordance with the strategic goals of WA Health" (WA Health, 2014).

Other organizations focus on data access by including researchers in their governance decisions. Statistics Canada has a long-standing network of subject matter advisory committees, most of whose members are academic researchers, and the ministerial-appointed National Statistics Council has substantial researcher membership (StatCan, 2014i). The CIHI Board of Directors also has traditionally included at least one prominent researcher. The leading international genetic research consortia, such as the International Cancer Genome Consortium (see Box 5.6 and 5.11), exist primarily for research purposes, and are led and run by leading researchers.

5.1.2 Managing Risk

As discussed in Section 3.4, the possession and use of health and health-related data involve a number of risks, first and foremost the risk that individuals' identities and sensitive information could be revealed. No data custodian or data steward can promise to eliminate all risks, as privacy risk cannot be completely eliminated. However, it can be significantly mitigated through effective risk management incorporating systematic processes for assessing and dealing with threats. The same is true for other risks involved, including risks of discrimination against or stigmatization of citizens, as well as risk to professional standing and reputation for those holding data if they grant access in a situation that could bring their organization into disrepute.

The International Organization for Standardization (ISO) lays out guidelines for risk management (ISO, n.d.), which can be adapted to the circumstances of any type of business or government agency. Box 5.3 describes how, for example, this standard has been adapted by Public Safety Canada for the Government of Canada, and augmented by the Information and Privacy Commissioner of Ontario to cover privacy issues. Based on these standards, risk management in providing access to detailed individual-level data, including linked data, is aimed at enabling access to data while recognizing the risk of privacy loss. The process of analyzing risks depends on the type of data released to researchers, as discussed in Chapter 2.

5.1.3 Respecting Privacy

Preventing inappropriate use of personal data is a key concern for the public. Some data custodians may choose to protect privacy by not allowing access to data at all, but thereby fail to act in the public interest by enabling appropriate access to the data. If data are to be shared, the core challenge is to enable access to data while maintaining appropriate protection of confidentiality.

Concern over data custodians' use of data in the NHS in England and Wales led to the release, in 1997, of the *Caldicott Report* on how patient information was controlled (DH, 1997). The objective was to ensure that confidentiality was not undermined by those handling data in the NHS. These concerns were recently re-examined and a subsequent report released in 2013 with the objective to ensure "an approximate balance between the protection of patient information and the use and sharing of information to improve patient care" (DH, 2013).

The key principles of the *Second Caldicott Report* (DH, 2013) are outlined in Box 5.4. The Panel believes these principles show how respect for privacy can be implemented in practice for data custodians, while also meeting other objectives. The addition of the seventh Caldicott principle in the second report

highlights the importance of enabling appropriate data sharing as a key objective of data custodians. The Panel recognizes that translating these principles into practice can be challenging given the overwhelming volume of data. In some cases, for example, exploratory data analysis may be a more optimal approach than conventional hypothesis testing, with its typically stringent requirement that reasons for performing research be strongly justified.

Box 5.3
Risk Management Steps: Canadian Examples

Public Safety Canada has articulated the following steps for managing risk in the public sector:

Step 1: Setting the Context: articulating an institution's objectives and defining its external and internal parameters to be taken into consideration when managing risk.

Step 2: Risk Identification: defining, recognizing, and recording risks.

Step 3: Risk Analysis: understanding the nature and level of risk, in terms of its impacts and likelihood.

Step 4: Risk Evaluation: comparing the results of risk analysis with risk criteria to determine whether a risk and/or its magnitude are acceptable or tolerable.

Step 5: Risk Treatment: identifying and recommending risk control or risk treatment options.

(PSC, 2012)

Further elements of risk management in the context of privacy have been added by the Information and Privacy Commissioner of Ontario:

Reporting: laying out policies on access, and establishing proactive notification of breaches, what data are accessed, how often, and by whom.

Monitoring for Compliance: external audit of risk control; periodic review of risk management policies, procedures, and operations by internal staff; and reviews by external and independent experts.

(IPCO, 2010)

Box 5.4

Caldicott Principles for the Protection of Personal Confidential Data

1. **"Justify the purpose(s):** Every proposed use or transfer of personal confidential data within or from an organization should be clearly defined, scrutinized and documented, with continuing uses regularly reviewed by an appropriate guardian.
2. **Don't use personal confidential data unless it is absolutely necessary:** Personal confidential data items should not be included unless it is essential for the specified purpose(s) of that flow. The need for patients to be identified should be considered at each stage of satisfying the purpose(s).
3. **Use the minimum necessary personal confidential data:** Where use of personal confidential data is considered to be essential, the inclusion of each individual item of data should be considered and justified so that the minimum amount of personal confidential data is transferred or accessible as is necessary for a given function to be carried out.
4. **Access to personal confidential data should be on a strict need-to-know basis:** Only those individuals who need access to personal confidential data should have access to it, and they should only have access to the data items that they need to see. This may mean introducing access controls or splitting data flows where one data flow is used for several purposes.
5. **Everyone with access to personal confidential data should be aware of their responsibilities:** Action should be taken to ensure that those handling personal confidential data — both clinical and non-clinical staff — are made fully aware of their responsibilities and obligations to respect patient confidentiality.
6. **Comply with the law:** Every use of personal confidential data must be lawful. Someone in each organization handling personal confidential data should be responsible for ensuring that the organization complies with legal requirements.
7. **The duty to share information can be as important as the duty to protect patient confidentiality.** Health and social care professionals should have the confidence to share information in the best interests of their patients within the framework set out by these principles. They should be supported by the policies of their employers, regulators and professional bodies."

(DH, 2013)

5.1.4 Maintaining Public Trust by Providing Evidence of Trustworthiness

A strong and growing body of evidence suggests that many members of the public are willing to bear some level of risk of a loss of privacy in order to contribute to research that has a reasonable prospect of delivering public benefit or wider social value (Haddow *et al.*, 2007; O'Doherty & Burgess, 2008; Dixon-Woods & Tarrant, 2009; Ipsos MORI, 2014). However, the evidence also indicates that tolerance of such risks is by no means unconditional; for example, common public concerns that must be addressed involve access by commercial entities to citizens' data and ensuring that there is tangible social value from access. This suggests that who has access to the data and how the resulting research findings are used remain important considerations for many members of the public. This trust in public research needs to be fostered and maintained if research is to continue. Moreover, this trust has to be respected, as individuals have a continuing interest in their data. Trust can be maintained and strengthened through transparency of governance, including providing evidence of the steps taken to protect privacy and of the benefits to the public from research. Furthermore, the principle of maintaining trust has implications for some of the other principles outlined above. It suggests, for example, that the nature and degree of risks to privacy must be weighed against the nature and likelihood of public benefit. These considerations directly affect risk management. In finding the appropriate balance of considerations across all four principles in any given context, trust will be maximized.

In most (if not all) cases, data have been collected by trusted professionals, whether by individual doctors and hospitals, academic researchers, or Statistics Canada. These data custodians preserve trust with the public by establishing and imposing rigorous policies aimed at preventing inappropriate access to data. This relationship cannot be compromised, and indeed must be safeguarded through ensuring that access to data is appropriate, and seen to be so. Continuing involvement of data custodians to ensure that decisions on accessing data are appropriate is warranted, as discussed later in this chapter.

The consequences of failing to secure and maintain the trust of citizens are profound. Loss of trust could have a direct and negative impact on citizens' care. It is well recognized in law and by professional bodies, for example, that there is a public interest in maintaining trust, because its loss could lead citizens to stop providing vital information about themselves that can be used in their own care and protection. This has clear implications for public health. From a research perspective, availability of data could be imperilled if public trust were to decline. As a result, maintenance of public trust is at the heart of effective governance.

While there are no specific established grounds for trust, Baroness Onora O'Neill, a philosopher and former chair of the Nuffield (U.K.) Council on Bioethics, argues that trust can be built in two steps:

> If we want others to trust us, the first step is to be trustworthy […] The second step is to show that we are trustworthy: we have to provide enough intelligible evidence of competence, honesty and reliability in the relevant matter for others to reach an intelligent judgment. This is not best done by showing that we have ticked all the prescribed boxes, kept perfect records or excelled in some league table. Complex forms of accountability may be useful for third parties, but what matters for most people in judging where to place their trust is generally simpler. Most of us look for evidence of trustworthiness — of competence, honesty and reliability — in relevant matters.
>
> (O'Neill, 2013)

Evidence of trustworthiness can be built up over time through such measures as:

- demonstrating practically that the four guiding principles common to the best practice entities are being respected;
- engaging proactively with the public;
- introducing and demonstrating best practices in governance;
- communicating the value of research output; and
- describing to the public the effort undertaken to protect individual information.

For example, MCHP has an award-winning (from CIHR) public communication program of short (four page) briefs on their research, which are readily accessible and often the subject of media stories. Statistics Canada and CIHI are in the news regularly with valuable new information. These are the results of explicit policies to develop a public profile and engender trust.

5.2 CROSS-CUTTING ASPECTS OF GOVERNANCE

As described earlier, the six best practice entities share with the other organizations within their respective jurisdiction's complex environment a *collective* responsibility for addressing four cross-cutting aspects of governance: privacy, research, information, and network governance. The separate treatment of the four overarching aspects of governance provided in this chapter addresses *comprehensive* governance; depending on the entity and circumstances, certain aspects will be more prominent. For example, REBs may be more concerned with research governance, while data custodians may be more concerned with information governance. The four cross-cutting aspects are outlined in Box 5.5.

While the four aspects of governance apply at all three levels of access (i.e., within a single jurisdiction, across Canada, and internationally), the application of each type of governance will differ at each level. In this chapter, much of the discussion of privacy, research, and information governance addresses best governance practices in single jurisdictions identified in the literature. The Panel did, however, also explore the types of governance mechanisms that have been put in place by organizations that are accessing and sharing data — across Canada and internationally — by forming networks. The discussion of such organizations will constitute part of the treatment of network governance below.

Box 5.5
Four Cross-Cutting Aspects of Governance

Privacy Governance: is concerned with how the "entities" (i.e., formal and informal organizations as well as individuals) in the complex environment of health and health-related data collectively address risks to privacy posed by the movement of data. Governance in this area requires specialized knowledge of technology, law, and statistical methods.

Research Governance: is concerned with how the entities collectively ensure that (i) the research that is undertaken is of high quality; (ii) there are significant benefits of research to society; and (iii) these benefits outweigh any risks, from practical, ethical, and legal perspectives.

Information Governance: is concerned with how the entities collectively manage information in a way that maintains confidentiality of the data while also providing appropriate access.

Network Governance: is concerned with how the entities collectively manage the research enterprise and share research results through the formation of networks. It involves creating common standards for data collection and developing policies for international data sharing. This is the broadest of the four overarching aspects of governance, since it can extend across multiple jurisdictions and complex environments, as well as involve entities from both the health and non-health sectors. Furthermore, it can involve consideration of how the other three overarching aspects of governance interact to create an effective platform for the undertaking of research.

Each of the four cross-cutting aspects of governance will now be explored. A comprehensive analysis would require addressing how each aspect pertains to each entity within the health and health-related data complex environment in Canada. For the sake of brevity, this chapter focuses on the entities within which each cross-cutting aspect of governance is especially relevant. To emphasize relationships between aspects of governance or different entities, sections of this chapter will be cross-referenced.

5.2.1 Privacy Governance

Respect for privacy is a requirement for each organization as well as each individual involved in the complex health and health-related data environment. It is reflected in the methods used to frame, collect, store, and provide access to data. Establishing that such processes are defined and implemented is key to privacy governance and requires expertise in technology, statistical methods, ethics, and privacy law to determine whether privacy is respected. Academic literature suggests that some REBs can be inconsistent in interpreting ethical and legal guidelines on what is identifiable information (Gershon & Tu, 2008; Willison *et al.*, 2008; Caulfield *et al.*, 2011; Yiannakoulias, 2011). For example, not all REBs include members with knowledge of what constitutes "reasonable" de-identification measures. To address the concern that specialized knowledge is needed to protect privacy, some countries and Canadian provinces have either established an external entity to ensure that privacy and confidentiality are respected or developed an information governance review panel dedicated to privacy review. An example of the former is MCHP (Manitoba), where a separate approval process of the Health Information Privacy Committee (HIPC) works in parallel with the REB (MCHP, 2014d). An example of the latter is SAIL (Wales).

Other examples of best practices in privacy governance are the following:

- **BORN (Ontario):** Data accessed from the BORN database in Ontario are certified to indicate that they have undergone de-identification and that the de-identification process has been approved by and reported to the REB (BORN, 2015).
- **ICES (Ontario):** Internal procedures determine differential access by individuals to data sets that have different degrees of de-identification. These policies have been developed in consultation with the Information and Privacy Commissioner of Ontario. REB notification is provided for health system evaluation projects conducted under Section 45 of PHIPA in Ontario. REB approval is sought for all research projects conducted under Section 44 of PHIPA (ICES, 2014b).

- **Statistics Netherlands:** Explicit criteria and processes (see Section 2.6) have been defined that flow from governing legislation for determining eligible institutions, eligible researchers within those institutions, and eligible research projects to be undertaken by the researchers (CBS, 2014a).

There are advantages and disadvantages to each of these approaches, and introducing one approach may lead to changes to the governance in other parts of the complex environment. For example, when SAIL enabled online data linkage, it introduced the separation principle as an added safeguard (discussed in Section 5.2.3). By contrast, when ICES was formed in 1992, its data holdings were restricted to its own researchers; consequently, ICES did not have to consider the risks from releasing data to a wider audience. Certification that data are de-identified in the BORN database may be sufficient for a single database.

Best Practice: The entities highlighted by the Panel have developed dedicated processes to evaluate privacy concerns when enabling data access.

5.2.2 Research Governance

While research governance entails many aspects, the panel chose to focus on the REB process. To comply with the ethics codes outlined in Chapter 4, university-affiliated researchers submit requests for approval to conduct data-based research to REBs and also apply to data custodians for access to the data. Since research using health and health-related data may give rise to a wide range of ethical concerns beyond privacy protection, such as stigmatization of groups or potential future discrimination, ethical review of such research is an important element of the research governance framework. In its review of the evidence on the effect of research governance on timely and appropriate access to data, the Panel found that timeframes for these approvals vary widely among organizations and jurisdictions in Canada, ranging from months to years. Research involving data or researchers from multiple sites increase the number of REB reviews involved, and linkage or data transfer between Canadian provinces can lead to further confusion and proliferation of REB approvals required, sometimes stymieing the research plan. This concern has led to some action in Canada; for instance, Alberta has reduced the number of REBs in the province from six to three (AIHS, 2013).

One other model is provided by New Zealand, which reduced the number of REBs for the entire country to four after reforms in 2012.[18] In keeping with good governance (Box 5.1), its new system is based on the principles of robustness, efficiency, transparency, and consistency. Standard operating procedures call for a single online application form that leads to a response within 35 calendar days, or 15 days for an expedited review (with a possibility of suspending the process once for up to 90 days). All four REBs can review multiregional projects, so that there is no duplication. Comprehensive minutes are published electronically to promote transparency and consistency. There is a process of appeal within a timeframe of 20 working days. Ethical review is kept separate from scientific review to prevent overlap with the peer review process (HDEC, 2012).

Vaughan *et al.* (2012) compared the New Zealand system with Australia's (which is similar to Canada's) in a case study of a research project on the epidemiology of serious conditions of pregnancy over the period 2009 to 2011. In New Zealand, researchers submitted a single application for an expedited review, and ethical approval across all 24 sites was granted. The entire ethics process required an estimated 10 hours of work by the researchers. In Australia, "as of September 2011, 46 full/expedited ethics applications, 131 site governance applications and 136 letters of support requests were made over 33 months, involving an estimated 3,261 hours by [investigators], and an associated resource burden by participating sites, to obtain approval to receive non-identifiable data from 291 hospitals" (Vaughan *et al.*, 2012). Hard-copy applications were required by 38 REBs. Of these, 26 REBs required multiple (range 2 to 27) hard copies of the application, resulting in over 17,000 printed pages (Vaughan *et al.*, 2012).

Best Practice: To minimize the number of approvals when performing cross-subject or cross-jurisdictional research — and therefore to improve timeliness — certain jurisdictions, such as Alberta, New Zealand, and Wales, have harmonized the REB process.

5.2.3 Information Governance

While privacy governance is concerned with the risks to privacy and research governance is concerned with the conduct of high-quality and ethical research, information governance is concerned with how the entities within the complex environment collectively handle information. This includes how they provide access, as well as under what circumstances and to whom they provide access, and how they address the legal, ethical, and quality standards relating to the

18 New Zealand had begun to reform its system following a parliamentary report in 2004. The number of REBs was reduced from 15 to seven: six regional boards and one REB responsible for all projects that involved more than one region (Jenkin *et al.*, 2006).

handling of sensitive and personal information. Although the term "governance" suggests that one is dealing with higher-level principles of some sort, "information governance" tends to be more concerned with operational matters of roles and practices, including those of individuals as well as organizations. Accordingly, best practices in this context will tend to be concrete illustrations of how to address concerns relating to how information is managed.

In part, information governance is concerned with how custodians strike an appropriate balance between ensuring confidentiality of data and enabling appropriate use of data. It does this by specifying requirements and standards to handle data "legally, securely, efficiently and effectively, and in a manner that maintains public trust" (DH, 2013). Requirements and standards may also apply to those who supply information (such as health-care professionals and institutions) to data custodians. Information governance defines who has the right to make decisions about information and who is accountable to ensure appropriate behaviour with respect to information. It also includes the processes, roles, policies, standards, and measures (such as quality assurance and audit) to ensure the effective and efficient use of information (Logan, 2010).

Although the scope of information governance encompasses the entire life cycle of data, the Panel chose to pay particular attention to only a subset of concerns, specifically the matters of data access, privacy management, de-identification, data safeguards, and risk management. Each is discussed below, with a primary focus on best practices as illustrated by the Panel's selected entities or by other organizations that enable access within a reasonable timeframe.

Data Access

How data are accessed has an effect on the risk of information being revealed inappropriately. The range of data access regimes is covered in Section 2.5. These different data access arrangements involve different risks and necessitate different governance arrangements.

In addition, there is public pressure to increase transparency of data, especially data held by public institutions (McNutt, 2015). This pressure is also fostered by the increasing tendency to make data available on the internet. As a result, there is a growing movement internationally towards "open data." However, it is impossible to control how data are used when put entirely in the public domain, through the internet or other media. Moreover, public release of data is not appropriate when there are privacy implications. Accordingly, even institutions committed to an open data agenda must consider appropriate information governance arrangements, as discussed in Box 5.6.

Box 5.6
Data Access Governance in International Research

As an example of how information governance can be put into practice, the International Cancer Genome Consortium (ICGC) established the International Data Access Committee (IDAC) to:

- develop policies for investigators to obtain access to controlled data;
- provide oversight to any ICGC member projects that are responsible for reviewing requests for such data; and,
- monitor compliance by bodies authorized to distribute ICGC data, and users of the controlled data.

The IDAC has broad geographic representation and includes representatives of the ICGC executive; experts in ethics, databases, and international law; cancer survivors; potential users of the data; and other independent lay persons — ideally, fewer than 20 members.

"The IDAC also develop[s] guidelines for practical approaches to providing qualified investigators with access to controlled data. In doing so, it consider[s] mechanisms and tools […] already in use by other organizations that distribute controlled data sets to international scientists […]. Potential users and their institutions are required to submit Assurance Agreement forms that include:

- a written description of the purpose of the research;
- an agreement not to try to identify or contact the donor subjects;
- an agreement not to redistribute controlled access data; and,
- plans to destroy controlled access data sets once they are no longer being used."

(ICGC, 2012)

Privacy Management as Part of Data Access

Inevitably, some individuals working with data — whether as data custodians, stewards, or researchers — will come into contact with individual-level data that are sensitive or can be linked with data from other databases, placing the confidentiality of information at risk. To encourage respect for privacy and to build a culture of respect for privacy, several of Canada's privacy officers have advanced privacy management programs (Table 5.1). Such a program assigns responsibility for privacy to certain positions in an organization and

encourages training of staff. A Canadian example is the privacy management program implemented at Population Data BC (Hertzman *et al.*, 2012); as well, privacy management is consistent with an approach called "privacy by design" advocated by the Information and Privacy Commissioner of Ontario (PbD, n.d.). In other organizations, privacy management extends as well to researchers, via agreements between data custodians and those requesting data access.

Note that privacy *management* is concerned with processes within an organization dealing with ensuring confidentiality. It is therefore more operationally focused than privacy *governance*, which is mainly policy oriented. Nonetheless the two are closely related, and some overlap is to be expected.

Best Practice: Certain entities have developed comprehensive and enforceable researcher-custodian agreements, such as Nova Scotia's Personal Health Information Act, to ensure that researchers maintain the confidentiality of the information that they receive.

Table 5.1

Building Blocks of a Privacy Management Program

Organizational Commitment	**a) Senior management support**
	b) Privacy Officer Responsible for the development and implementation of the program controls and their ongoing assessment and revision; role and responsibility for monitoring compliance are clearly identified and communicated throughout the organization.
	c) Privacy Office Supports the ability of staff to monitor compliance and fosters a culture of privacy within the organization.
	d) Reporting
Program Controls	**a) Personal information inventory** Identifies the personal information in the entity's custody or control; its authority for the collection, use and disclosure of the personal information; and the sensitivity of the personal information.
	b) Policies On collection, use and disclosure of personal information, which include requirements for consent and notification; access to and correction of personal information; retention and disposal of personal information
	c) Risk assessment tools **d) Training and education requirements** **e) Breach and incident management response protocols** **f) Service provider management** **g) External communication**

Adapted with permission: OPCC *et al.* (2012)

Appropriate Institutional Structure and Separation Principle

Linking data sets across organizations might raise the possibility that a large number of individuals employed across several data custodians could access large amounts of identifiable data. However, institutional structures can be established to minimize the risk to confidentiality when linkage is performed. One approach for managing employee access is to separate identifying data and content data. Separation can be achieved in different ways, but the specific methods used by different organizations all follow the *separation principle*. In general terms, this principle means that any given individual cannot see both the identifiable data used to link data sets (e.g., name, address, date of birth) and the content data (e.g., clinical information, benefit information) (NSS, n.d.-a).

The separation principle can be observed by using an external organization (typically called a trusted third party) to deal with identifying information or by managing all data internally but ensuring that identifying data and content data are administratively — and sometimes physically — separated. SAIL, Data Linkage WA, and Statistics Canada all use some form of the separation principle (Data Linkage WA, 2014a; SAIL, 2014; StatCan, 2014e).

SAIL is an example of an organization that uses an external trusted third party. Data custodians divide their data into a demographic component (containing name, address, gender, date of birth, etc.) and a content component. They assign a *join key* (also called a *linkage key*) to each part of the data. The demographic component is sent to the National Health Service Wales Informatics Service (NWIS) and the content component is sent directly to SAIL. NWIS replaces identifying data with a unique, encrypted code, and sends this code (along with minimal demographic information on area of residence, week of birth and gender) to SAIL. The two components of the data set are then re-combined at SAIL using the join key and made available for researchers to access, subject to approvals (SAIL, 2014).

In contrast, at Statistics Canada, both the data custodian and trusted third party roles are internal, but the processes involve essentially similar separation. For example, the details of the record linkages involved in creating and accessing the Social Domain Record Linkage Environment are spelled out in a lengthy Privacy Impact Assessment (PIA) signed by the Chief Statistician and submitted by Statistics Canada to the Office of the Federal Privacy Commissioner. This PIA includes detailed descriptions of the data sets involved, the flows of various kinds of data among different processing steps, a threat risk assessment, and enumeration of compliance with the ten prescribed privacy principles (StatCan, 2014e).

Best Practice: Entities that use the separation principle for data linkage, such as SAIL, Statistics Canada, and Data Linkage WA, have minimized the risk of inadvertent access to data with nominal identifiers by staff.

De-Identification of Data

Because a precise standard for de-identification of data is lacking, laws across Canada and in many countries have generally called for a "reasonable" method to be followed (see Section 4.3.3). The hallmark of such a process is a well-documented risk-based approach with objective risk measures. It follows generally accepted statistical and scientific principles, and is documented and transparent (El Emam & Malin, 2014).

As part of information governance, an organization needs to follow and document a de-identification process meeting the requirements of a "reasonable" process of de-identification that would involve the following principles and process (Appendix B contains a high-level elaboration of how this is done in practice):

1. Identify and classify variables in the data
2. Mask the direct identifiers
3. Determine the threshold for de-identification
4. De-identify data
5. Report on certification

A key part of a risk management strategy is to determine the objective risk thresholds. De-identification steps can be adjusted to meet a pre-set re-identification risk; as discussed below, the risk of re-identification can be kept as low as warranted.

However, de-identification can result in a loss of accuracy or usefulness of the data. There are well-developed methods for evaluating the impact on data quality from de-identification (for example, measures of information loss) (El Emam *et al.*, 2009). There is also evidence that when sophisticated de-identification methods are used such information loss can be minimized, in some cases with negligible impact on the conclusions of the data analysis (CCO CPO & CIO, 2011). The degree of such loss can be adjusted depending on who uses the data and in what circumstances. The degree of de-identification can be lessened and the usefulness of the data improved if, for example, researchers receiving the data (i) have a track record of using data for research, (ii) sign confidentiality contracts, and (iii) hold data in secure locations (as discussed above).

Best Practice: Robust de-identification techniques that meet legal standards (i.e., de-identification is "reasonable") make it possible to reduce the risk of re-identification to a level that is appropriate for a given access mode (and its accompanying security controls). Best practice includes ensuring that de-identification is documented, transparent, and meets statistical thresholds for re-identification risk while maintaining data utility.

Technology's Role in Enabling Access to and Safeguarding Data

As well as lowering the cost of generating and analyzing data, technology can also play a role in safeguarding data. A key concern, as outlined in Chapter 4, is that data on individuals can flow out of the jurisdiction in which the data custodians reside and in which contracts are signed, and data custodians may have no means of enforcing appropriate use in another jurisdiction.

This traditional model of data physically flowing to the researchers who then undertake the analysis locally — wherever they are located — can be reversed, so that the data continue to reside with the data custodian, but the analysis moves physically to the data. The most obvious ways in which this happens are when researchers undertake their analysis in a secure safe haven under the control of the data custodian. Many Canadian organizations, including CIHI, MCHP, ICES, and Statistics Canada, have had such "safe havens" for many years. When safe havens are used, the data never leave their home jurisdiction and principal legal responsibility and accountability can continue to rest with the relevant agency.

Another approach where the analysis moves to the data involves providing a researcher with a secure connection to a remote computer that is under the control of the data custodian and on which both the sensitive data and the researcher's analytical software reside. The researcher then runs the software remotely and only obtains the results for his/her later use, not the original data. For example, the Population Health Research Network (PHRN) in Australia has developed the Secure Unified Research Environment (SURE), a computing environment that allows researchers to access the approved data extracts for their research project and analyze them remotely. Researchers see a facsimile of the screen of the remote virtual computer on their local computer screen. "Within the SURE, each researcher is allocated a virtual computer that runs entirely on hardware physically located at and controlled by the SURE. The SURE also contains extensive data storage and back-up capabilities and a range of analysis software for researchers to use" (PHRN, 2014).

This model cannot work, however, if data across multiple jurisdictions are required to be pooled. In this case, technology holds the prospect of helping as well. The traditional model of data flowing to researcher can also be broken

up into its component pieces. For example, Wolfson *et al.* (2010) propose an approach — called DataSHIELD — in which the core identifiable data remain at each data custodian, and summary statistics that enable statistical analysis flow freely via the internet. Critically, the summary statistics should not contain identifiable information, or information that could be used to generate identifiable information. Others suggest further ensuring confidentiality protection by encrypting the raw data (El Emam *et al.*, 2012, 2013).

Such technologies allow data analysis to be conducted across borders while the original custodians in each jurisdiction retain principal responsibility for their portion of the data. These new methods offer a means to transcend current barriers to data flows between provinces. They also offer a means of developing national systems of linkable data while keeping the data with the data custodian. A different kind of example of the effective use of such a pan-jurisdictional approach is CNODES, described in Box 3.6. However, all such solutions involve a high level of knowledge and skills in software design and implementation. Consequently, for these solutions to be adopted, data custodians would need to learn about these new technologies, engage suitably skilled staff, and develop appropriate governance systems to ensure that the information flowing via the internet does not contain identifiable information, and that the code written to enable operation of such systems is rigorously checked.

While technological means to facilitate safe and secure data access and use are invaluable, they are not a complete solution to good governance challenges. All decisions about data access and the attendant arrangements also involve careful ethical judgments. In assessing acceptable levels of risk, decision-makers need to keep in mind that no technical measure is 100% risk-free.

Best Practice: Certain entities successfully maintain data confidentiality through safe havens or encrypted access or both. These entities include Statistics Canada and Statistics Netherlands. Key features of a well-functioning safe haven include mechanisms to approve researchers, robust internal and external monitoring and oversight, and continual review of governance arrangements.

Best Practice: Technology is continuously changing. New technologies can be adopted and developed to improve the safeguards on confidentiality. Given the central importance of technology in this area, it is critical to have individuals with knowledge of its importance involved in governance.

An additional factor to consider is the economics of the different options for data access such as de-identification, access at secure "safe haven" centres, and remote access. Some cost-benefit analyses have been done for certain approaches (i.e., de-identification) (El Emam, 2013c); however, without economic

comparisons of the different practices (or data to make these comparisons), an important yardstick for comparing the various alternatives was unavailable to the Panel.

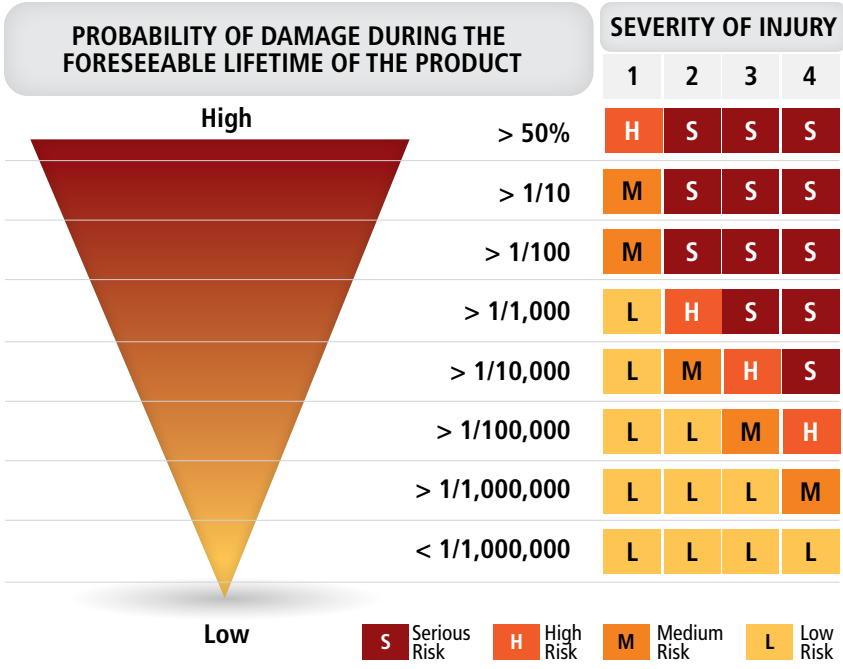
Establishing an Acceptable Level of Risk

A critical element of any information governance model is the determination of what level of risk is "acceptable." The TCPS defines *minimal risk research* as "research in which the probability and magnitude of possible harms implied by participation in the research is no greater than those encountered by participants in the aspects of their everyday life that relate to the research" (CIHR *et al.*, 2014). However, the "everyday life" standard, developed to deal with traditional types of research, may no longer be relevant, given that the risk of confidential data being revealed from online website visits and social media postings in everyday life is now quite high.

By contrast, the American Society of Safety Engineers defines *acceptable risk* as "the probability of a hazard-related incident or exposure occurring and the severity of harm or damage that could result" are as low as reasonably practicable, and tolerable in the setting considered (ASSE, 2011). The *as low as reasonably practicable* standard is the "level of risk which can be further lowered only by an increment in resource expenditure that cannot be justified by the resulting decrement of risk" (ASSE, 2011). This definition reflects key concepts: (i) some level of risk has to be tolerated (risk cannot be completely eliminated), (ii) this level of risk is more than minimal, and (iii) ensuring a minimal level of risk could entail unreasonable expenditure.[19]

Several benchmarks to conceptualize the level of risk have been developed around the world and across several industries, ranging from nuclear safety to pharmaceuticals. The European Commission has guidelines for product safety and has developed a risk assessment matrix (Figure 5.1). This matrix was designed to conceptualize a traditional type of risk (injury), and may not be directly relevant to modern, information technology risks such as data breaches, but is offered as an example of such an approach. In the matrix, the severity of injury is graded from level 1 (such as a mild allergic reaction or temporary pain in the eye that does not require treatment) to level 4 (such as fatality or permanent loss of sight). As the nature and severity of harm increase, the level of risk that is defined as low declines (European Commission, 2010). To give some indication of the probabilities in Figure 5.1, the draft guideline labelled

19 A similar definition used at the European Environment Agency, for example, states that the acceptable risk level is the "level of risk judged to be outweighed by corresponding benefits or one that is of such a degree that it is considered to pose minimal potential for adverse effects" (EEA, n.d.).



Reproduced with permission: European Commission (2010) http://eur-lex.europa.eu, © European Union, 1998–2014.

Figure 5.1

European Commission Risk Assessment Matrix for Product Safety, Based on the Combination of Injury Severity and Probability

Under the European Commission risk assessment matrix for product safety, risk is "the combination of the severity of possible damage to the consumer and the probability that this damage should occur" (European Commission, 2010). An injury can vary in severity and each injury scenario happens with a certain probability. Thus, if minor injuries (e.g., level 1) caused by the use of a product are highly likely to occur (e.g., > 50%), then the product may still be labelled as high risk. In contrast, if the product has the potential to cause severe injuries but these injuries are extremely unlikely to occur, then it can still be considered a low risk product.

probability between 1 in 1,000 (0.1%) and 1 in 10,000 (0.01%) as conceivable, but highly unlikely, and between 1 in 10,000 (0.01%) and 1 in 100,000 (0.001%) as practically impossible (European Commission, 2010). A re-identification attack simulated to evaluate the vulnerability of data in the United States (reviewed in El Emam *et al.* (2011b)) found that an estimated 0.013% of records could be re-identified correctly, which would be considered conceivable, but highly unlikely, and on the borderline of practically impossible using the European Commission matrix. Furthermore, an "adversary" who is willing to spend the time and resources required for an attempted re-identification attack would presumably see some value in the data. However, as discussed in Section 3.4.1,

the literature indicates that most "adversaries" are actually researchers who are attempting to evaluate whether a re-identification risk exists, rather than individuals who are hoping to use the re-identified information (El Emam *et al.*, 2011b).

The system in Scotland adopted a different approach called "proportionate governance" in which the level of scrutiny given a proposal depends on the level of risk (see Box 5.7). This framework also embraced a proactive, iterative approach involving data custodians, research communities, and public engagement so that the core objectives of all stakeholders were respected. The resulting Good Governance Framework (GGF), initially developed by SHIP, consisted of four elements:

1. An account of responsibilities of key actors and decision-makers (largely a matter of clarifying who is a "data controller" under European data protection law);[20]
2. A training facility for researchers leading to SHIP accreditation (delivered through distance learning);
3. A statement of Principles and Best Practices to guide decision-making; and
4. A mechanism based on principled proportionate governance to assess data linkage requests (based on actual risks and calibrated accordingly).

(Laurie & Sethi, 2012)

Box 5.7
Proportionate Approach to Governance in Scotland

The Scottish Informatics Programme (SHIP), funded by the Wellcome Trust from 2009 to 2013, was a collaborative interdisciplinary initiative between Scottish universities and NHS Scotland, the principal data custodian of much of Scotland's medical data, including hospital admissions and discharges as well as cancer registries. SHIP's objective was to build on the long-established, high-quality data sets held within Scotland to maximize the value of research using these data sets, to reduce undue regulatory burden, and to maintain appropriate ethical scrutiny. Core features of the model were subsequently adopted by NHS Scotland and the Scottish government. The work now continues across sectors beyond health, and across the United Kingdom, as part of the Farr Institute (SHIP, n.d.-b).

20 Available at the Scottish Informatics Programme website (SHIP, n.d.-a).

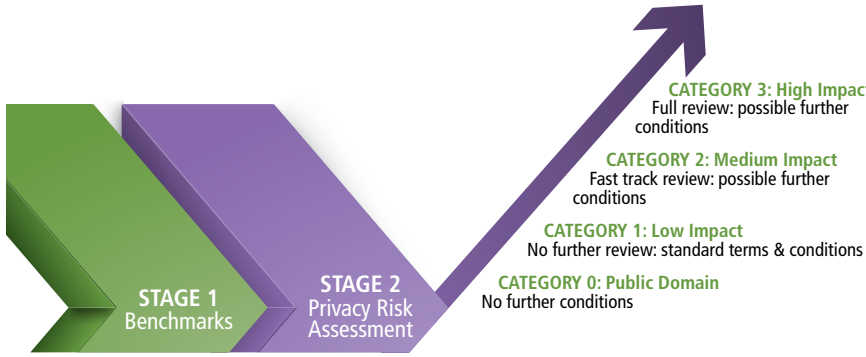
The GGF, which is now being built upon by the Farr Institute @ Scotland, embodied a principles-based approach that recognized that sensitive decisions must be made with careful exercise of judgment. Accordingly, the governance framework was not overly prescriptive, setting hard-and-fast rules. Rather, it provided decision-makers with key principles to consider when scrutinizing data linkage requests. Key principles included demonstrating reasonable prospects of delivering benefits in the public interest through data linkage, and identifying and developing management strategies for privacy risks. If consent for linkage was not granted, the refusal needed to be fully justified. This last feature of the GGF was a direct result of findings from public engagement exercises (Laurie & Sethi, 2012).

A crucial, additional pragmatic component of the GGF was its mechanism for delivering risk-based proportionate governance. It asked three fundamental questions about each linkage request:

> **Safe people:** Were researchers accredited by SHIP or an equivalent body?
>
> **Safe data:** Were data linked through SHIP and its indexing and linkage service?
>
> **Safe environments:** Were researchers using a SHIP safe haven or equivalent?
>
> (Laurie & Sethi, 2012)

If the answer to all of these questions was "yes," a fast-track route was available for researchers, requiring no further scrutiny of the application. If the answer to any of these questions was "no," further scrutiny was taken, including full review by an ethics body, such as the Privacy Advisory Committee in Scotland. Additionally, a privacy risk assessment ensured that the levels of governance and oversight addressed the relative risks of any proposal, including those meeting all three benchmarks (Laurie & Sethi, 2012). Figure 5.2 shows this approach.

The proportionate governance mechanism has been adopted within key bodies responsible for data linkages within NHS Scotland (NSS PAC, 2013). The Principles and Best Practices statement has also served as the basis of a public consultation on cross-sector data linkage undertaken by the Scottish government (The Scottish Government, 2012).



Reproduced with permission: SHIP (2012). This work was supported by the Wellcome Trust through the Scottish Informatics Programme (SHIP) Grant (Ref WT086113). SHIP was a collaboration between the Universities of Aberdeen, Dundee, Edinburgh, Glasgow and St Andrews and the Information Services Division of NHS Scotland.

Figure 5.2

Proportionate Governance at SHIP: Categorization of Data Access Applications According to Stratified Categories of Risk

The categorization approach developed by SHIP was designed to make the process of application review faster and more efficient for researchers and data custodians. The risk categories "correspond directly to increasingly stringent terms and conditions that must be met in order to achieve authorization for a linkage to go ahead" (Sethi & Laurie, 2013). Category 0 relates to data already in the public domain. Category 1 involves applications where risks are thought to be minimal or negligible. Category 2 applications have issues that might be flagged for possible further consideration. Category 3 applications would be subjected to full Privacy Advisory Committee approval mechanisms.

Best Practices: The European Commission has developed a systematic method for characterizing risk. Scotland has integrated a proportionate approach to risk in its governance system.

Practical Application of Information Governance

Box 5.8 summarizes professional standards established for accessing health-related data in the United Kingdom. It provides an example of policies for implementation of information governance in practice, respecting the guiding principles highlighted by the Panel.

Box 5.8
Professional Standards and Good Practice for Accredited Safe Havens in the United Kingdom

In the United Kingdom the following requirements for data stewardship must be met by accredited safe havens:

- attribution of explicit responsibility for authorizing and overseeing the anonymization process (e.g., through a Senior Information Risk Officer);
- appropriate techniques for de-identification of data, the use of "privacy enhancing technologies" and re-identification risk management;
- use of "fair processing notices." These notices, a requirement of European data protection legislation, are sent to data subjects to inform them that personal data are being processed for stated purposes;
- a published register of data flowing into or out of the safe haven, including a register of all data sets held;
- robust governance arrangements including, at a minimum, policies on ethics, technical competence, publication, limited disclosure/access, regular review process, and a business continuity plan including disaster recovery;
- clear conditions for hosting researchers and other investigators who wish to use the safe haven;
- clear operations control, including human resources procedures for information governance, use of role-based access controls, confidentiality clauses in job descriptions, effective education and training, and contracts;
- a standard for information security commensurate with ISO 27001 (the ISO standard for Information Security Management Systems) and the NHS's Information Governance Toolkit;
- clear policies for the proportionate use of data, including competency at undertaking privacy impact assessments and risk and benefit analysis;
- standards that are auditable;
- a standard template for data sharing agreements and other contracts that conforms to legal and statutory processes;
- appropriate knowledge management, including awareness of any changes in the law and a joint approach with others working in the same domain; and
- explicit standard timescales for keeping data sets, including those that have been linked, which should be able to support both cohort studies and simple "one-off" requests for linkage.

(DH, 2013)

While access to linked administrative, registry, and vital statistics data is available in some Canadian provinces, these arrangements generally cover data for broad groups of people rather than specific subgroups. However, sometimes questions apply to more specific populations. For instance, is the prevalence of diabetes rising quickly among First Nations populations? What are special health needs of immigrant children? Are patients living with HIV developing other chronic diseases? Addressing these questions may require linkage of, for example, provincial data to special identifiers held by other authorities such as the federal government or public health departments. Linking these data requires specific data sharing agreements between organizations that can legally disclose and receive such data. The information created by these linkages is sensitive and warrants that representatives of the communities be closely involved in the governance of the data sets. This highlights an important point: information governance within an entity might not cover the range of considerations at stake when sharing data across organizations, provinces, or countries. For such reasons, it is important also to consider the wider notion of network governance that can, among other things, include a range of collaborations among stakeholders.

5.2.4 Network Governance

The creation of collaborative research networks, perhaps involving not just a circle of researchers but also other stakeholders such as data custodians and funding agencies, has the potential to maximize social benefits flowing from data-oriented research. Indeed, networks focusing on particular areas of research have been formed both nationally (Level 2) and internationally (Level 3).

Among the benefits of building a research network is that it may be the only way to amass enough data to conduct a study. For example, international collaborations have made it possible to discover genetic variations that are associated with susceptibility to conditions such as cancer. A by-product of such international collaborations is that, since the data involved need to be comparable, definitions and standards must be defined in advance. Indeed, since a single network-based study may involve researchers from more than 100 organizations (Milne *et al.*, 2014), networks are a central contributor to standardization and harmonization (discussed in Chapter 2 and Box 5.9). Sharing of best practice at the network level can also improve governance (Di Iorio *et al.*, 2013).

While the advantages of networks are many, they also have some pitfalls. For example, unless care is taken in establishing a large network, it may be unclear who is responsible for coordinating the network to ensure that it provides the benefits envisaged. Without clear lines of responsibility, networks can fail to

reach their objectives. In addition, the creation and continued nurturing of networks is time-consuming and expensive, especially so for the researchers involved, some of whom are forced to spend an increasing fraction of their time on network management rather than research.

Box 5.9
Coordinating Data Collection for Comparable Results

Data standards are a key element in facilitating data-based research, and numerous examples demonstrate how standards can be put in place by national or international networks. In addition to research networks, which are formed to address common research goals, another type of network involves the boards and committees of major organizations. These networks are composed of individuals who may have different research interests and diverse professional backgrounds, but who share the common goal of developing national or international standards.

Standardization has been a core function of the WHO and CIHI. The WHO has developed a family of international classifications (FIC) to provide a "consensual, meaningful and useful framework which governments, providers and consumers can use as a common language" (WHO, 2015a). The International Classification of Diseases (ICD) is one of its three main classifications, which cover basic health parameters. Other more specific classifications, such as the ICD for Oncology, build on these main classifications (WHO, 2015a). The ICD-10 is used by all WHO Member States and has been translated into 43 languages. It is currently under revision, with an expected release date of 2017 for the ICD-11 (WHO, 2015c). The revision steering group consists of medical specialists; researchers; experts in terminology, classification, and public health; and WHO-FIC representatives. The group meets regularly to ensure that the combined expertise of its members is used to produce a consistent and coherent document that addresses user needs (WHO, 2015b).

In Canada, CIHI supports international standards for use by Canadian health-care professionals. For example, it tailored the ICD-10 to meet Canadian morbidity needs by producing the ICD-10-CA. CIHI has produced numerous standards to promote data quality and consistency in databases that keep track of information on hospital visits (CIHI, 2014g). Its Board of Directors links federal, provincial, and territorial governments with non-governmental health groups and is composed of Deputy Ministers of health, CEOs of hospitals and health networks, and other health-care leaders (CIHI, 2014f).

Best Practice: Certain entities, such as the WHO and CIHI, have put standards in place before data collection to enable prospective data harmonization.

Another challenge confronted by multi-jurisdiction networks involves differing implementations and interpretations of privacy and other laws, which can hinder the pooling of data (this matter was also discussed in Chapter 2). Arising from this consideration, the Panel identified two types of multi-jurisdiction networks: those that share individual-level data, and those that share only aggregated data. Examples of the former include multiple genomics initiatives, such as the database of Genotypes and Phenotypes (dbGaP) and the ICGC (see Boxes 5.6, 5.10, and 5.11). Examples of the latter include the FDA's Sentinel System, CNODES, and EUBIROD (see Section 2.4.3). It is important to note that, whereas pooling of individual-level data may be ideal, valuable results can still be generated by networks that only share aggregated data.

Best Practice: When it is not possible to pool individual-level data, other models, such as CNODES in Canada and DataShield in Europe, have been successful in enabling statistical analysis across jurisdictions.

It is also important that networks develop standardized data security protocols. This task is simplified, although still challenging, when all members of the network are governed by a central directive for protection of personal data. This is the case for networks that only involve participants from the European Union (e.g., EUBIROD), which has developed a single data protection directive for all member states (Di Iorio *et al.*, 2009). Genetics initiatives such as the dbGaP (Box 5.10) are among the most advanced in their successful development of policies for international sharing of individual-level data. An additional example is the ICGC (Box 5.11), which includes members from Canada, the United States, China, Japan, the United Kingdom, Germany, Mexico, India, Australia, and more.

Box 5.10
The Database of Genotypes and Phenotypes

In the United States, the National Institutes of Health (NIH) recognized the fundamental importance of making the rapidly growing volume of genetic and phenotypic research data more broadly available for secondary use. To this end, the NIH has established the database of Genotypes and Phenotypes (dbGaP). This is a highly secure repository for individual-level data from a growing number of different studies. But at least as importantly, it is a set of security and policy protocols and metadata holdings for each of these many studies (NCBI, 2015).

Before researchers receive NIH funding for a large-scale study involving human genomic data, they must provide a detailed plan on how they will make their data available to dbGaP or another NIH-designated repository; thus, sharing data is a pre-condition for funding (NIH, 2014). These genomic data sharing plans must include information on planned data formatting, the specific respondent consents that will be asked in the study, the ethics approval process, and details on how data elements will be described (NIH, 2007). Once the study is complete and all the data have been collected and cleaned, they are deposited with dbGaP.

For secondary users, dbGaP provides two levels of access — open and controlled. Open access (publicly available) data are limited to non-confidential information. But for typical secondary analysis, researchers need "controlled access" to phenotype and genotype data for individual participants (NCBI, 2015). To gain such access, the researcher must submit an application that includes details of the planned data analysis, the names of collaborators, the physical and computer security setup, and whether the proposed research is consistent with the consents provided to the original study (NCBI, 2013). An important control mechanism is the organization of data by *consent groups*, which "consist of all of the data from study participants who have agreed to the same data use as specified in the informed consent for the study" (NCBI, 2012). Thus, to receive a complete data set containing all participants in a study, a researcher must apply for data from all consent groups and use them only as permitted.

In sum, the NIH dbGaP is an extensive, carefully developed set of protocols and data holdings that strikes the fundamental balance between protecting the confidentiality of study participants' data and maximizing the public benefits from research using those data.

Box 5.11
The International Cancer Genome Consortium

The ICGC, a large-scale international initiative launched in 2008 to analyze 25,000 cancer genomes, recognized at the beginning stage of its creation that a framework had to be developed to govern international sharing of the data its members would generate (ICGC, 2014). Its foundational policy, "Goals, Structures, Policies and Guidelines," set forth a comprehensive list of data that would be deposited in either open or controlled access categories (ICGC, 2008). The list was understood to be dynamic, so that certain types of data could be transferred from the open to the controlled category and vice versa in response to changing scientific, technical, legal, and ethical issues.

ICGC also created a controlled access mechanism that regulates access to certain sensitive data (e.g., detailed phenotype and outcome data, genome sequence files) by requiring third parties to apply to a data access committee (which itself is overseen by an international data access committee) and complete a simple but authoritative data access form that contains privacy safeguards. This controlled data access mechanism is seen as a best practice to both protect study participants from re-identification and data misuse by third parties and allow researchers streamlined and user-friendly access to useful but sensitive personal data that are crucial for biomedical research studies (Joly *et al.*, 2012).

Best Practice: When legal systems differ, methods have been developed to further research by multinational consortia such as dbGaP and ICGC.

Scientific Integrity

In the context of data network governance, the Panel uses the term *scientific integrity* to refer to management of a range of practices that can affect the accuracy of research findings. These include the quality and the completeness of the data and the quality of the methods used to analyze the data. Very large data sets are particularly problematic, not only because they were not designed primarily for research, but also because it is possible to select multiple outcome variables and analytical approaches and selectively report only those that support a particular hypothesis. This is more likely in observational research, with an exploratory component than in randomized trials where most outcomes and analyses are pre-specified. Concerns have been expressed about the number of published research findings that prove to be inaccurate (Ioannidis, 2005). This potential for bias and error increases with the degree of statistical complexity and the number of variables and analyses. Inadequate statistical analysis of data

creates a significant risk of invalid results for a wide body of research across a broad swath of subjects. Developing best practices in data curation, pre-specification of variables and statistical analyses, and registration of sufficiently detailed protocols are important roles for research networks. Furthermore, as one of the key tenets of science is replicability, making data readily available to facilitate replication will become increasingly necessary to scientific progress.

In analysis of large data sets, risk of inaccuracy arises because of the increased complexity of the data, with anywhere from dozens to thousands of dimensions captured. Not only does effective analysis of such databases test computational power and storage capacity, but what is called "high-dimensional" modelling also challenges standard statistical methodologies because of statistical problems associated with the volume and complexity of the data (Fan *et al.*, 2011; Einav & Levin, 2013; Fan *et al.*, 2014). Failing to undertake adequate statistical controls may lead to results that are not meaningful, and, in the health context, that may create harm by leading to incorrect conclusions that are then applied to care (for example, if the estimated effects are inaccurate).

A network of the research community can mitigate these potential problems with scientific integrity by putting in place standards for statistical analysis, sharing information on issues in high-dimensional modelling, and acting to address any incorrect conclusions — whether through research misconduct or faulty analysis — released publicly. These are fundamental issues that merit substantial further thought, but were considered to be beyond the mandate of the Panel.

5.2.5 Informing Governance through Public Engagement

Informed support from members of the public for research conducted with their data requires that they are aware of the research that is taking place, they approve of the policies and methods that are being used, they trust the researchers and data custodians who are handling their information, and they understand the benefits that may result from this research. These requirements can only be met if the public is engaged in the research process. The lessons learned from public engagement have the potential to inform all four aspects of governance. Although much of the work on public engagement has been in the experimental research or primary care setting, more recent work has explored public involvement in studies using existing health and social data (McKenzie & Hanley, 2007; Born & Laupacis, 2012; Ipsos MORI, 2014; Jones *et al.*, 2014).

For example, CFHI (2014) highlights how Alberta Health Services introduced patient engagement research, which involves training citizens living with various health conditions to design and conduct qualitative health research. They then collaborate with health professionals and researchers and engage other patients in research.

Another form of patient engagement is patient-centred outcomes research, which focuses on outcomes of importance to patients, such as quality of life. This approach promises to enhance decision-makers' ability to fully understand and weigh alternatives. In Canada, one objective of CIHR's Strategy for Patient-Oriented Research, for instance, is to ensure that the right patient receives the right intervention at the right time. This involves bringing innovative diagnostic and therapeutic approaches to the point of care, to ensure greater quality, accountability, and accessibility of care (CIHR, 2013a). Engaging with patients concerning their experiences of care and the likely benefits of research provides an opportunity for patients to understand how research benefits them directly as well as the population at large.

Both types of public engagement discussed above — in which members of the public are involved in deciding whether or how to conduct research and in understanding how research benefits them — are at the level of individual projects. However, public engagement can be at the level of research in general. McKenzie and Hanley (2007) discuss how public engagement can play a role at many stages, to improve uptake of research benefits and strengthen support and understanding of research. In fact, many researchers in teaching hospitals also promote research as a positive benefit to the community.

To support a project-level or more general engagement framework, CIHR (2013c) has recently provided a set of principles. Working with citizens has the potential to add value to research programs or projects, while mutual learning and understanding build trust and credibility. Openness enhances transparency and accountability. Public engagement should be inclusive in its approach, supporting individuals to ensure their full participation. In addition, engagement should support communities. In Canada, there has been public engagement with specific cultural or socio-economic communities. For example, research is conducted in Aboriginal communities through the First Nations Framework, which involves those communities in research. Research networks working with Aboriginal data sets must make particular efforts to engage appropriately with the communities and their representative organizations. It is essential to build a relationship of trust in which the Aboriginal communities know that their interests are being protected. In Canada it is important that researchers comply with the Tri-Council policy on the Ethics of Health Research

Involving First Nations, Inuit and Métis People (CIHR, 2013b). As discussed in Section 3.4.3, ICES has established an arrangement with First Nations in Ontario by entering into a data governance agreement that incorporates, among other elements, the OCAP (Ownership, Control, Access and Possession) principles (First Nations Centre, 2007; Antone *et al.*, 2014).

Public engagement also has the potential to improve public understanding of privacy risks and protection in research involving linked data. A recent report on public attitudes towards research uses of administrative data in general carried out in the United Kingdom by Ipsos MORI found that those surveyed were broadly happy with data linking if (i) research has social value, (ii) data are de-identified, (iii) data are secure, and (iv) businesses cannot access the data for profit (Ipsos MORI, 2014). The report concluded that further work is needed to understand fully the public's view on commercial involvement. These findings reflect similar results from public engagement within Canada's health sector.

Examples of best practice consistently show that public engagement can be incorporated into the governance framework of data linkage initiatives for research. In the SAIL system in Wales, for example, there is a specialized consumer panel to provide a public perspective on research based on linked data. Its role is to act as advisers on research issues and on engaging with the public, guide recruitment of the public to steering groups for studies, provide views on data protection issues, discuss proposals for research, review information designed for a lay audience, and act as advocates for research based on linked data (SAIL, 2013; Jones *et al.*, 2014). Representatives of the consumer panel also sit on the panel overseeing information governance at SAIL.

The Farr Institute @ Scotland has been designed with an integral program of public engagement. The program has found high levels of trust in public institutions that use data for public benefit, but it has also revealed evidence that members of the public also want to be able to exercise some degree of continuing control over their data. This finding was reflected in SHIP's GGF, which requires applicants for data linkage or use to address whether their research questions can be answered using data obtained with consent, and, if not, to justify use of data without consent in terms of a strong public interest (Sethi & Laurie, 2013).

In a guide written for health and medical research organizations, McKenzie and Hanley (2007) caution that researchers must work with community members to decide precisely how value can be added to research through community participation. Although public engagement has great potential,

measuring its impact is challenging. Much of the current evidence on impact centres on qualitative studies with relatively small focus groups, and more robust measurements of impact are needed (Barber *et al.*, 2011a, 2011b; Purtell & Wyatt, 2011).

5.3 ALLOCATING RESPONSIBILITIES FOR ASPECTS OF GOVERNANCE

Without a clear allocation of roles and governance responsibilities among the entities within Canada's complex environment dealing with the provision of timely access to data, there is risk of overlap, duplication, and confusion. In short, developing clear lines of responsibility in connection with the four overarching aspects of governance constitutes best practice. Without an integrated approach to determining roles and responsibilities in this sense, there is a clear risk of delay in accessing data, for example, since steps may be missed or inconsistent rulings may need to be sorted out. Clarifying the responsibilities of key entities in Canada's complex environment could be a positive step in enabling timely access to health and health-related data for research. Table 5.2 summarizes the roles of different groups (e.g., researchers, data custodians, policy makers) and governing bodies (e.g., REBs, privacy monitoring boards) in overseeing various aspects of governance and provides examples of entities that are following best practice by successfully performing these roles.

Table 5.2

Allocation of Governance Responsibilities

Stakeholder Group or Governing Body	Aspect of Governance				Role of Group or Governing Body and Examples of Entities Following Best Practice
	Privacy	Research	Information	Network	
Privacy Monitoring Board	✓				Dedicated board to provide oversight of policies at data custodians, and to interact with researchers to evaluate data requests for privacy concerns **Examples:** IGRP in Wales, HIPC in Manitoba
Research Ethics Board		✓			Dedicated board to provide review of ethical and legal implications of research while remaining aware and informed of risk to privacy **Examples:** IGRP in Wales, REB system in New Zealand Ensure that proposals to use data are appropriate and in the public interest **Example:** IGRP in Wales

continued on next page

Stakeholder Group or Governing Body	Aspect of Governance				Role of Group or Governing Body and Examples of Entities Following Best Practice
	Privacy	Research	Information	Network	
Researchers				✓	Participate in networks to ensure that research conducted is scientifically valid and ethical
			✓		Follow protocols to hold data securely
			✓		Be aware of the risks to confidentiality; follow data-confidentiality agreement **Example:** researcher-custodian agreements in Nova Scotia, MCHP
				✓	Participate in efforts to standardize and harmonize data **Examples:** various genetics consortia
		✓			Publish and archive all methodological details so their analysis can be replicated **Example:** Statistics Canada
Data custodians			✓		Implement Caldicott Principles for the Protection of Personal Confidential Data, including clarity that use of data is necessary
			✓		Develop clear mission statements, clarifying that the duty to share information can be as important as the duty to protect data confidentiality **Example:** ICES
			✓		Establish a privacy management program, such as training of staff on the importance of protecting privacy, establishing appropriate risk management policies, and establishing clear lines of responsibility **Example:** Population Data BC
				✓	Engage public on benefits of research **Examples:** Farr Institute @ Scotland, SAIL
	✓			✓	Participate in efforts to standardize and harmonize data **Examples:** CIHI, WHO, Statistics Canada Publish descriptions of all record linkages **Example:** Statistics Canada For large data projects, perform privacy impact assessments **Examples:** Statistics Canada, Canadian Health Measures Survey

continued on next page

Stakeholder Group or Governing Body	Aspect of Governance				Role of Group or Governing Body and Examples of Entities Following Best Practice
	Privacy	Research	Information	Network	
Data access office			✓		Authenticate the qualifications of researchers seeking access; verify security plan and institutional approval **Example:** SHIP good governance framework
Individuals conducting data linkage			✓		Ensure that identifiable data are held securely
			✓		Follow best-practice techniques for de-identification of data by establishing clear, transparent processes with objective measures of risk **Examples:** ICES, Farr Institute @ Scotland
			✓		Implement separation principle **Examples:** SAIL, Data Linkage WA,Statistics Canada
Health-care providers (such as doctors and nurses)			✓		Ensure accurate and complete data entry; obtain consent; harmonize data
			✓		Ensure accurate and complete data entry; obtain consent; harmonize data
				✓	Engage public on benefits of research **Examples:** Alberta Health Services, CIHR Strategy for Patient-Oriented Research
Senior policy makers				✓	Ensure that value of data holdings is maximized through enabling access for the public benefit; set framework for data access; communicate public benefit of research **Examples:** Government of Manitoba, Western Australia Health Department

Among the international (but still Level 1) best practices chosen by the Panel, the case of Wales (Box 5.12) illustrates a governance innovation introduced to make the overall system for accessing data for research more efficient. The new system created a single agency, the Independent Governance Review Panel (IGRP), to oversee health data access for de-identified data. The main thrust of the innovation was to ensure that the risks of re-identification were minimized while access to data was made rapid and efficient. Note that, since data are de-identified, individual consent is not required, but ethical review is maintained to ensure that there are no ethical risks to segments of the population.

Box 5.12
Governance Integration in Wales

In Wales, health data are held in linkable form at a university (SAIL at Swansea University), but the linkage is conducted by NHS Wales, as a trusted third party. As a government agency, NHS Wales is trusted to hold identifiable information such as names, addresses, postal codes, etc. Control over access rests with a single independent panel, the IGRP, in which NHS Wales participates. The IGRP ensures that the data go through appropriate de-identification, thus addressing the ethical concern over the risks to privacy. To ensure other ethical concerns are also respected, the IGRP includes representatives from ethics review as well. The IGRP comprises representatives from the British Medical Association, the National Research Ethics Service (ethics review), Public Health Wales (policy advice), NHS Wales Informatics Service (data custodian and trusted third party), and the SAIL Consumer Panel (SAIL, 2014).

5.4 CONCLUSION

The entities dealing with the provision of access to data for research purposes in Canada are best thought of as a "complex environment," the parts of which were not designed explicitly to work in concert with one another as an integrated system with a common overall purpose and a coordinated set of roles and responsibilities. Compared with best practices from other jurisdictions where such an integrated approach has been taken (needless to say, with due regard to the overarching issues of privacy, information, research, and network governance), there is a risk of overlap, duplication, and confusion, and hence delays in accessing data. Nonetheless, over time, coordination, consistency, and overall effectiveness of Canada's complex environment can be improved through the adaptation of the pre-existing entities with an eye to the Canadian as well as international best practices outlined in this chapter (that is, unless the responsible government(s) see(s) fit to undertake a broad review of organizational roles and alignment comparable to that undertaken in Wales or Scotland).

This chapter has provided examples of governance best practices, recognizing that the precise form of governance adopted by an organization or other entity depends on the individual organization, its mandate, the constraints within which it operates, the types of data involved, and other considerations. Governance of data is in a process of evolution, and the sharing of experiences and practices through networks will strengthen governance in future.

6

Conclusion

- Main Charge
- Overcoming Technological and Methodological Challenges of Integrating Data
- Benefits, Risks, and Barriers to Timely Access to Data
- Legal and Ethical Considerations
- Best Practices for Governance to Improve Access while Maintaining Confidentiality
- Final Thoughts

6 Conclusion

Addressing the main charge and the five related sub-questions engaged the Panel in a study of considerable breadth as well as depth. Indeed, much of the content of this report is likely to be of greatest interest and value to experts in specific areas, such as statistics/analytics, law/ethics, and policy/governance.

6.1 MAIN CHARGE

The ***main charge*** to the Panel posed the following question:

> ***What is the current state of knowledge surrounding timely access to health and social data for health research and health system innovation in Canada?***

The Panel's response can be briefly summarized as follows.

Providing health researchers and health system innovators with timely access to health and social data is important, and likely to grow more important in the future. The volume of data that could be used by these communities for the benefit of Canadians is expanding rapidly; taking advantage of these data represents a substantial opportunity. Indeed, it can be argued that there is an ethical imperative to use these data for the public good. Survey data suggest that Canadians generally support the use of their health-care encounter and related data, including evolving EHRs, for research.

Providing timely access to such data presents special challenges:

- One challenge is purely technical. The sheer range of different types of data involved, and the number of individuals and organizations collecting and holding the data, means that the development of common standards and formats is extremely complicated.
- A second arises because the range of organizations, programs, and communities involved in the provision of timely access to data in Canada, as in other jurisdictions, is very broad, and constitutes what this report has called a "complex environment of heterogeneous entities." In most places the environment is only loosely coordinated, with duplication of roles and in some instances outright conflicts between the goals of different entities.
- A third involves striking a balance between providing timely access to data and respecting privacy.

All jurisdictions have to grapple with these challenges. While there is no single, ideal definition of roles for the entities in any jurisdiction's "complex environment of heterogeneous entities," some have been more successful than others in this effort. The Panel has highlighted six entities that constitute best practice, three from Canada and three from other countries.

The Panel's review of practices from these six entities, as well as others, shows that improvement in the current state of affairs in Canada is possible. Barring the redesign of provincial, territorial or, indeed, the pan-Canadian complex environment, these improvements could come from the adoption of some of the best practices articulated in this report, and the review by the appropriate governing bodies of the governance lessons learned from these best practices.

The next four sections expand on this summary, and point to the Panel's primary responses to the sub-questions raised in its charge. Although there is a rough parallel between the sub-questions and the chapters of this report, with the bulk of the Panel's findings on a given question found in its corresponding chapter (with two addressed together in Chapter 5), there are also overlaps.

6.2 OVERCOMING TECHNOLOGICAL AND METHODOLOGICAL CHALLENGES OF INTEGRATING DATA

There are substantial benefits to be obtained from data linkage and pooling; in Chapters 2 and 3, several examples were given where research has been advanced in this way and public benefits have been delivered. There are also serious challenges involved in integrating data: some have to do with the data themselves (e.g., consistency of standards across studies); others have to do with legal/ethical constraints (e.g., prohibition against pooling data from different jurisdictions).

The first sub-question raised directly the matter of how to deal with these challenges:

Sub-Question 1: *What is known about how to address technological and methodological challenges associated with linkage of health and social data from various sources and across jurisdictions?*

As noted above, Chapter 2 reviewed many of the challenges associated with data integration in detail. The Panel found it useful to adopt a tri-level distinction for clarifying the different concepts and challenges involved in timely access to data (recall Figure 1.1). Level 1 involves access to one or more data sets all

within one province or territory; Level 2 involves access to data sets spanning multiple provinces or territories; and Level 3 describes access to data sets collected or residing in multiple countries. Some issues apply primarily to one or two levels of access and others span all three. The Panel arrived at several findings that directly spoke to the challenges of data integration. Potential solutions are provided for the various challenges.

6.2.1 Harmonization

Challenges

Individual-level data held in different databases are more easily compared if they are collected using common standards and definitions; otherwise, retrospective harmonization is required to make them comparable. Harmonization is necessary for all forms of data integration at Levels 1, 2, and 3. Certain domains of information are more difficult to harmonize than others. Furthermore, it can be challenging to ensure that the precise meaning of any given data element is as close as possible across the different data sets being combined. Harmonization is particularly problematic for unstructured data such as free-form text.

Solutions

The most durable solution for harmonization is to develop standard terminologies, questionnaires, measurements, and operating procedures before data are gathered. Although such standardization is the ideal situation, other approaches can be considered if this approach is too challenging, time-consuming, labour-intensive, or underlying consensus is absent.

If the data were collected without standards in place, retrospective harmonization could help make use of the existing information. For example, tools are available to map one set of clinical diagnostic codes to another, and techniques such as DataSHaPER can help determine whether similar inferences can be drawn from variables across different studies. However, the potential to harmonize existing information is necessarily always limited by the heterogeneity of the data collected and may not be possible for some types of information.

For unstructured data, a more pre-emptive method is to collect the information in a structured format in the first place using a standard nomenclature such as ICD-10. A retrospective (but second best) solution is to use natural language processing to convert textual files into codified terms or tags drawn from controlled vocabularies.

6.2.2 Linkage

Challenges

Data linkage allows different types of information for one individual to be brought together. Because an individual, for the most part, receives health and health-related services in a single province or territory, linkage typically requires access to data from multiple organizations in one jurisdiction (Level 1 access). It can be challenging if data have been strongly de-identified or if unique identifiers are available for some individuals in a data set but not others.

Solutions

Even without unique identifiers for each individual, probabilistic methods can be used to link records. If data need to be linked, the simplest way, if possible, is to link prior to de-identification. Databases do not always need to be linked permanently to undertake most research. The link can be destroyed after the research is completed, and/or kept completely separate by a trusted third party.

6.2.3 Pooling

Challenges

Data pooling is often used to increase sample size by pooling similar data from several populations. It can occur within a province or territory, but is often most relevant to multi-jurisdictional access (Levels 2 and 3) because data from different jurisdictions will likely be stored in separate databases. In numerous important cases, restrictive interpretations of privacy and other laws within Canada are hindering interprovincial pooling of individual-level data. International data pooling is even more difficult, especially for health-care records.

Solutions

Various approaches have been developed to avoid the need to pool individual-level data. For example, summary data meta-analysis, used by CNODES involves harmonization of individual-level data across different studies followed by statistical analyses at each study site, and, finally, pooling of the (non-confidential) summary statistics to obtain an overall result. Another solution, provided by DataSHIELD, uses sophisticated iterative techniques to mimic a pooled analysis of data from individual participants when, in reality, the data always remain with their original data custodian. A further extension of the DataSHIELD approach involves encryption of the individual-level data on each host computer.

6.2.4 Access

Methods for providing researchers access to data vary from secure physical locations, to secure online links for approved researchers, to publicly available aggregated data. Along this spectrum, the degree of risk to privacy is lowered, as the information shifts from only mildly de-identified data made available to researchers under strict and secure conditions to more strongly de-identified individual-level data to aggregated data or data analysis results.

Current rules and procedures, to authorize research and to allow data access, overlap and are often time-consuming. Furthermore, processes and requirements for access are sometimes unclear. Depending on the jurisdiction, delays may be caused by slow approval from REBs, other governing bodies, or data custodians, or by incomplete applications from researchers. Decisions about access are not always consistent. As a result, the ability to access and link data for a given research project within reasonable timeframes is uncertain, uneven across Canada, or even lacking. Solutions to these access challenges are discussed in Section 6.5.

6.3 BENEFITS, RISKS, AND BARRIERS TO TIMELY ACCESS TO DATA

Sub-question 2: *What is known about the benefits, risks and barriers to timely access to health and social data for health research and health system innovation in Canada?*

The Panel responded to this sub-question in Chapter 3, the main findings from which are summarized below. The matter of better communication with the public on benefits and risks was addressed again in Chapter 5 in the context of the imperative of moving from a culture of caution to a culture of trust (see Section 6.4).

Benefits

There are major benefits to increasing the appropriate use of individual-level health and health-related data to improve patient care, facilitate innovation, and generally improve the health of Canadians. Most obvious is more timely identification of adverse drug reactions.

Risks

Based on evidence from business and government, the main privacy risks of working with identifiable data are deliberate (e.g., malicious attack) or inadvertent (e.g., human error) data release. Although health data breaches can cause serious harm, the risk of a breach actually occurring in the context

of research is low, particularly if effective governance mechanisms are in place and respected by care providers, researchers, and data custodians. Calibrating various degrees of de-identification of individual-level data with corresponding degrees of access security is another effective strategy for managing privacy risks.

An additional risk arises from research on specific communities, which has the potential to make individuals within these communities feel stigmatized. This risk can be appropriately mitigated by involving communities in the research process.

Barriers

Numerous barriers can impede health data research, including logistical and ethical barriers to achieving access and issues with the data themselves, such as lack of comparability between data sets.

Organizations are also sometimes hesitant to share data, and there are numerous reasons for this reluctance. In the context of many competing priorities, data custodians may be reluctant to undertake the considerable effort required for data preparation, particularly if they do not have an adequate budget and/or a specific mandate to support research uses of their data. Anecdotal evidence from the Panel also revealed an understandable reticence of some organizations to enable research access to data that might reveal poor performance. Fear of lawsuits due to unclear privacy and ethics laws may result in incorrect or overly conservative interpretations of legislation, thereby impeding data sharing.

An additional barrier involves the potential for research to be hindered by lack of public engagement. A lack of communication with the public about their opinions on the use of personal data for research can lead to overly restrictive interpretations of the law. Informing the public of the controls that have been implemented to safeguard their information, as well as the benefits that have resulted from research using health data, has the potential to increase public confidence in and enthusiasm for this type of research.

6.4 LEGAL AND ETHICAL CONSIDERATIONS

Sub-question 3: *What are the ethical, legal, and social implications of timely access to such data?*

Chapter 4 addresses this question at length. Some of the findings of that exploration are presented here.

There are ethical imperatives to protect the confidentiality of individuals' data, on the one hand, and to provide access to quality data that enable research in the public interest, on the other. These two imperatives need not conflict, particularly when data do not include identifiable information or when data contain identifiable information and appropriate safeguards are in place.

Data custodians have fundamental legal duties to protect confidentiality of personal data, and these duties underpin their conduct. These duties can lead to cautious and conservative interpretations of allowable access when a complementary mandate to enable access to data for research is not made explicit.

Canadian federal and provincial/territorial laws generally address identifiable information and do not constrain researchers' access to de-identified or non-identifiable information. However, given the imprecise and inconsistent definitions of the term "identifiable information" in laws and ethical guidelines from different jurisdictions, it is difficult to be sure whether a data set qualifies as non-identifiable. Instead, it is useful to view de-identification as a continuum and to adjust access controls accordingly to mitigate re-identification risk.

Canada's governance of research ethics is fragmented, with significant differences across the provinces/territories. As well, laws on sharing data across provinces/territories and between countries differ or are lacking, sometimes leading to confusion for researchers and REBs about whether, or on what basis, data can be shared.

While participant consent is a cornerstone of experimental research involving humans, the ethical and legal considerations for accessing personal information are not the same as those for physical involvement in research. There may be sound ethical reasons to pursue research with health and health-related data without consent in some circumstances, notably when risk is managed and the research benefits the public good. Appropriate risk management involves keeping measures to protect privacy proportional to the potential harms of proposed research.

Among these findings, one deserves special emphasis. The ethical and legal frameworks currently in place in Canada appear to strike a balance between enabling research and respecting privacy. In practice, however, lack of knowledge and trust has led to a conservative implementation of these frameworks. Indeed, evidence suggests that current data use is too often subject to undue restrictions that inhibit timely access and adversely affect health and social outcomes. As discussed above, striking the appropriate balance between the public benefits of research and the risk of a loss of confidentiality is an important challenge for all jurisdictions. If Canada is to move from a "culture of caution," which arguably prevails in many organizations dealing with health data today, to a "culture of trust," Canadians must be engaged in the dialogue.

6.5 BEST PRACTICES FOR GOVERNANCE TO IMPROVE ACCESS WHILE MAINTAINING CONFIDENTIALITY

Sub-question 4: *What are best practices for improving access to such data for researchers while ensuring appropriate privacy safeguards and also taking full advantage of the digital data revolution?*

Sub-question 5: *What are best practices in Canada and internationally for governance frameworks that facilitate access to such data and maintain public trust in the research enterprise?*

The Panel chose to answer the last two sub-questions together in Chapter 5 because of their joint reliance on the identification of best practice organizations that provide timely access to health and social data. Six entities were highlighted. Although from different jurisdictions, their operations shared four underlying principles: (i) enable appropriate use of data, (ii) manage risk, (iii) respect privacy, and (iv) maintain public trust by providing evidence of trustworthiness. In themselves, adoption of these principles by entities involved in the provision of access to data can be regarded as an element of best practice.

In addition, the international "best practice entities" reviewed by the Panel operate within systems where a deliberate effort has been made to create a governance framework that addresses four cross-cutting aspects of governance: privacy, information, research, and network governance. Within their respective jurisdictions, this conscious effort to address overall governance in such a systematic way led to these entities sharing many of the following "earmarks" of best practice; not all earmarks will or need be found in every situation, but each is worthy of consideration in the Canadian context:

- dedicated processes (parallel to REBs) that specifically evaluate privacy concerns when enabling data access;

- maintenance of data confidentiality through safe havens and/or encrypted access;
- comprehensive and enforceable researcher-custodian agreements to ensure that researchers maintain the confidentiality of the information that they receive;
- use of the separation principle to minimize the risk of inadvertent access to data by staff;
- use of robust de-identification techniques to reduce the risk of re-identification to a level that is appropriate for a given access mode (and its accompanying security controls), with de-identification that is documented, transparent, and meets statistical thresholds;
- adoption of new technologies to improve the safeguards on confidentiality and lower the costs of conducting research;
- development of systematic methods for characterizing risk, and adoption of a "proportionate approach";
- minimization of the number of approvals needed when performing cross-subject or cross-jurisdictional research;
- harmonization of the REB process;
- adoption of standards prior to data collection to enable prospective data harmonization;
- adoption of other models, such as sharing of summary statistics across jurisdictions (e.g., the CNODES approach), when it is not possible to pool individual-level data;
- development of multinational consortia such as ICGC when legal systems differ; and
- engaging the public with evidence of the usefulness of the research and the steps being taken to protect privacy and confidentiality.

6.6 FINAL THOUGHTS

Effective use of health and health-related data — including social and administrative data — by the country's health researchers and system innovators will contribute to the development of beneficial health policies that improve the quality of health care for Canadians. This need will increase in the future as technology continues to develop and digitized data become more abundant.

However, timely access to health and health-related data for research in Canada varies across the country. While some jurisdictions have developed processes that provide access to data within a period of four months, others can take a year or longer. The reasons for these delays are multifold, such as concerns over data quality, lack of a roadmap for how to access data, fear of potential legal liabilities or embarrassing results, or media attention in the case of data breaches.

The Panel found that legal definitions and interpretations differ across provinces/territories and countries, which can lead to confusion or overly cautious interpretations of whether data can be accessed or shared, and that careful ethical judgments must be taken (sometimes in the absence of specific laws). However, the Panel also found that good governance can ensure the co-existence of access to data and respect for ethical principles and the law. Indeed, the prevailing legal structure in Canada calls for a system of governance to guide data custodians in their decisions on whether and how to give access to data. The Panel has identified several best practices, with a strong emphasis on public engagement, that provide the necessary guidance to help transform what is known as a culture of caution to a culture of trust.

The Panel concludes, therefore, that it is possible, in practice, to balance the competing responsibilities of protecting individuals' privacy, while simultaneously enabling timely access to data for the purposes of health research and health system innovation.

Glossary

Glossary

Appropriate Use: Appropriate use of data is enabled by laying out requirements and standards to handle data "legally, securely, efficiently, and effectively, and in a manner that maintains public trust" (DH, 2013).

Best Practice: Policies and practices currently in use by entities that collect, analyze, provide access to, and regulate laws surrounding access to, data, that — according to evidence identified by the Panel — are already helping to improve timely access while still protecting privacy. See also ***Good Practice***.

***Bona Fide* Researchers:** Researchers who "generate new knowledge and understanding using rigorous scientific methods," who intend to publish their research and share their data, and who conduct research in compliance with ethical and legal requirements as well as recognized good practice (MRC, n.d.).

Confidentiality: The duties and practices of people and organizations to ensure that individuals' personal information only flows from one entity to another according to legislated or otherwise broadly accepted norms and policies. In the context of health data, restrictions on and authorities for communicating personal information arise primarily from legislation, duties relating to professional obligations, or contracts. In these cases, confidentiality is breached whenever personal information is communicated that is not authorized by legislation, professional obligations, or under contractual duties.

Controlled Vocabulary: "A controlled vocabulary only includes terms that have been selected by the group that created the vocabulary. The goal of such a vocabulary is to standardize and simplify the organization of data and knowledge in a particular domain" (Kohane, 2011).

Data Custodians: Organizations that collect and/or hold data and make initial decisions on data use, disclosure, retention, and disposal. Data custodians play a central role in enabling or inhibiting access to health and health-related data by implementing policies on data collection, use, and disclosure. They also endeavour to ensure that their employees follow appropriate practices, such as keeping data secure. Data custodians include ministries of health, hospitals, primary care physicians, and regional health authorities.

Data Element/Data Record: In a set of data, a data element or "data field" is one attribute pertaining to an individual data record. A data record is a number of data elements pertaining to an individual or to an event like a health-care encounter. For example, a data element could be a date or a location.

Data Harmonization: The processes involved in producing inferentially equivalent data, thus allowing data from different sources to be compared or combined and used meaningfully in statistical analysis (e.g., regression).

Data Linkage: The process of "bringing together from two or more different sources, data that relate to the same individual, family, place or event" (Holman *et al.*, 2008). For example, linkage may be used to bring together information about an individual's health status, prescription drug use, and social media habits.

Data Pooling: The process of bringing multiple data sets together for analysis. Often, the data are from individual participants, rather than summarized results. In contrast to data linkage, which brings together different data pertaining to the same individual, data pooling brings together sets of similar data (ideally harmonized) from different individuals, so that a relationship amongst variables can be examined with a larger sample.

Data Standardization: The process of developing and implementing identical methods and tools for data collection, which, when used at different sites, will produce data that are already harmonized and thus comparable.

De-identification: The act of minimally perturbing individual-level data to decrease the probability of discovering an individual's identity (El Emam *et al.*, 2011b). It involves ***masking*** direct identifiers (e.g., name, phone number, address) as well as transforming indirect identifiers that could be used alone or in combination to re-identify an individual (e.g., birth dates, geographic details, dates of key events). If done correctly, de-identification is a defensible, repeatable, and auditable process that consistently provides assurance, based on generally accepted and repeatable statistical methodologies, that there is a very small risk of re-identification of any data that are released (El Emam, 2013b). The Panel chose to use the word de-identify rather than other similar terms in common usage that have various interpretations (such as *obfuscate* or *anonymize*).

Electronic Medical Record: "An electronic version of the paper record that doctors have traditionally maintained for their patients and which is typically only accessible within the facility or office that controls it" (CMPA, 2014).

Electronic Health Record: "A compilation of core [electronic] health data submitted by various healthcare providers and organizations, accessible by numerous authorized parties from a number of points of care, possibly even from different jurisdictions" (CMPA, 2014).

Good Practice: Policies and practices that — based on a combination of anecdotal evidence, literature review, and Panel analysis — have the potential to improve timely access while still protecting privacy. See also ***Best Practice***.

Governance: Governance can best be thought of as the role of an organization's board of directors or its equivalent that is focused on defining that organization's purpose and the development of the strategies, objectives, values, and policies that frame how that purpose will be pursued. It includes the development of such things as mission statements, statements of organizational objectives and values, logic models, organizational performance metrics, risk management frameworks, policies and guidelines for financial and operational matters, stakeholder relations, etc.

Harmonization: See ***Data Harmonization***.

Health Data: Data on health status of individuals (e.g., well-being, health conditions), health system performance (e.g., accessibility, effectiveness), and community and health system characteristics (e.g., resources) (CIHI, 2014h).

Health-Related Data : Data on non-medical determinants of health such as health behaviours, living and working conditions, personal resources, and environmental factors (CIHI, 2014h).

Identifiable Data: "Data is identifiable if the information contains the name of an individual, or other identifying items such as birth date, address or geocoding. Data will be identifiable if the information contains a unique personal identifier and the holder of the information also has the master list linking the identifiers to individuals. Data may also be identifiable because of the number of different pieces of information known about a particular individual. It may also be possible to ascertain the identity of individuals from aggregated data where there are very few individuals in a particular category. Identifiability is dependent on the amount of information held and also on the skills and technology of the holder" (European Commission, 1999; OECD, 2013d).

Information: The output of a process that analyzes, summarizes, interprets or otherwise represents data to convey meaning (Cabinet Office, 2012).

Information Governance: "How organizations manage the way information and data are handled within the health and social care system […]. It covers the collection, use, access and decommissioning as well as requirements and standards organizations and their suppliers need to achieve to fulfill the obligations that information is handled legally, securely, efficiently, effectively and in a manner which maintains public trust (UK Government, 2013).

Masking: Involves "the application of a set of data transformation techniques without any concern for the analytical utility of the data. This is a good approach for fields that are not required to be analyzed." Masking is applied to direct identifiers such as name and phone number. Masking techniques include, among others, removal of direct identifiers or replacement of direct identifiers with pseudonyms. In contrast to masking, statistical de-identification is applied to indirect identifiers and involves "the application of a set of data transformation techniques such that the resulting data retains a very high analytic value" (El Emam, 2013b).

Minimal Risk Research: "Research in which the probability and magnitude of possible harms implied by participation in the research is no greater than those encountered by participants in the aspects of their everyday life that relate to the research" (CIHR *et al.*, 2014).

Natural Language Processing (NLP): "A field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. NLP techniques allow the text in electronic medical records to be transformed from a clinical narrative to a set of codified terms or tags that are more readily subject to computational and statistical analysis" (Kohane, 2011).

Network Governance: How entities interact to manage research, so that it becomes the responsibility of the research community collectively rather than of individual organizations. Such governance includes, for example, establishing and publicizing common standards and definitions and developing policies for international data sharing.

Personal Health Data: "Personal health data encompass a wide range of information about an individual, which all touch upon an individual's private life. A health biography could include not only basic medical data: a history of all medical diagnoses, diseases and medical interventions, medications prescribed, test results, including imaging, etc. but could also include more

sensitive data: on mental health, relevant […] family history, behavioural patterns, sexual life, social and economic factors, etc. and health care administrative data: admissions and discharge data routine operational data, insurance and financial transactional data, etc." (European Commission, 1999).

Personal Health Identifier: A number, code or other element used in a health system to uniquely identify an individual, such as a health insurance number.

Pooling: See ***Data Pooling***.

Privacy: A broad concept that in Canadian law encompasses *personal privacy* (protection of one's physical self), *territorial privacy* (protection of one's private physical space), and *informational privacy* (protection of information about oneself and one's activities) (SCC, 2004).

Privacy Governance: In the context of this report, privacy governance monitors the risk to privacy posed by data requests from researchers, and the practices of data custodians in providing data (information governance) to ensure that confidentiality is protected. Such governance requires specialized knowledge of technology, law, and statistical methods.

Proportionate Governance: In the context of this report, proportionate governance refers to keeping the procedural mechanisms that researchers and data custodians must follow when engaged in data sharing and linkage proportional to the degree of risks associated with such practices. Proportionate governance operates in situations that are too variable to be regulated by hard laws (e.g., custom data access requests). It requires that analytical judgments be performed to ensure that the governance mechanisms deployed for a given research proposal correspond to the level of risk it entails (Sethi & Laurie, 2013). Proportionality is an important cross-cutting consideration across all types of governance that are put in place.

Research Governance: Among many other things, research governance ensures that the benefits to society of research outweigh any risks, from both an ethical and legal perspective.

Safe Haven: A physically, electronically, procedurally, and otherwise secure computer data system that *bona fide* researchers can access physically through an on-site visit or remotely through secure Internet connections (ISD Scotland, 2010b).

Social Data: See ***Health-Related Data***.

Structured Data: "Structured data are those that can be easily organised, stored and transferred in a defined data model, such as numbers/text set out in a table or relational database that have a consistent format (e.g. name, date of birth, address, gender, etc.)" (Kitchin, 2014).

Timely Access: Access granted and provided within a reasonable timeframe; in this assessment, access granted within four months of submission of a data request is considered timely access.

Unstructured Data: Unstructured data (e.g., free-form text) do not have a common identifiable structure. These data can often be searched as long as they are digital, but they are more difficult to use for computer analyses (Kitchin, 2014).

References

References

Acquisti, A. (2010). *The Economics of Personal Data and the Economics of Privacy: 30 Years After the OECD Privacy Guidelines*. Paper presented at Joint Working Party for Information Security and Privacy (WPISP) and Working Party on the Information Economy (WPIE) Roundtable, Paris, France.

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies, 42*(2), 249-274.

Adams, C. & Allen, J. (2014). Government databases and public health research: Facilitating access in the public interest. *Journal of Law and Medicine 21*(4), 957-972.

AIHS (Alberta Innovates Health Solutions). (2013). HIA-Designated REBs Reduced to Three and Full Reciprocity on its Way: Update on Health Research Ethics Harmonization. Retrieved September 2014, from http://www.aihealthsolutions.ca/news-and-events/media-centre/hia-designated-rebs-reduced-to-three-and-full-reciprocity-on-its-way/.

Al-Shahi, R., Vousden, C., & Warlow, C. (2005). Bias from requiring explicit consent from all participants in observational research: Prospective, population based study. *British Medical Journal, 331*(7522), 942. doi: 10.1136/bmj.38624.397569.68.

Allen, J., Holman, C., Meslin, E., & Stanley, F. (2013). Privacy protectionism and health information: Is there any redress for harms to health? *Journal of Law and Medicine, 21*(2), 473-485.

AMA (American Medical Association). (2012). *Crosswalking Between ICD-9 and ICD-10*. AMA.

AMS (Academy of Medical Sciences,). (2006). *Personal Data for Public Good: Using Health Information in Medical Research*. London, United Kingdom: AMS.

AMS (Academy of Medical Sciences). (2011). *A New Pathway for the Regulation and Governance of Health Research*. London, United Kingdom: AMS.

Anomaly, J. (2011). Public health and public goods. *Public Health Ethics, 4*(3), 251-259.

Antone, T., Henry, D., Jones, C., & Khan, S. (2014). *First Nations Health Data Linkage: A Collaborative Research Approach*. Paper presented at International Health Data Linkage Project, Vancouver (BC).

APEC (Asia-Pacific Economic Cooperation). (2014). APEC Cross-border Privacy Enforcement Arrangement (CPEA). Retrieved July 2014, from http://www.apec.org/Groups/Committee-on-Trade-and-Investment/Electronic-Commerce-Steering-Group/Cross-border-Privacy-Enforcement-Arrangement.aspx.

Arabsky, S., Alke, K., & Choi, T. (2014). Why does it take so long? Anatomy of a Data Request in BC. Vancouver (BC): Population Data BC, University of Brithsh Columbia.

ASSE (American Society of Safety Engineers). (2011). *ANSI/ASSE Z3590.3-2011, Prevention through Design: Guidelines for Addressing Occupational Hazards and Risks in Design and Redesign Processes*. Des Plaines (IL): ASSE.

Atkinson, G. (2014). Big data — What is it and what use is it? *Jorunal of Ambulatory Care Management, 37*(3), 196-198.

Austin, L. & Lemmens, T. (2009). Privacy, Consent, and Governance. In K. Dierickx & P. Borry (Eds.), *New Challenges for Biobanks: Ethics, Law and Governance*. Antwerp, Belgium: Intersentia.

Awadalla, P., Boileau, C., Payette, Y., Idaghdour, Y., Goulet, J. P., Knoppers, B.,... Laberge, C. (2013). Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *International Journal of Epidemiology, 42*(5), 1285-1299.

Barber, R., Beresford, P., Boote, J., Cooper, C., & Faulkner, A. (2011a). Evaluating the impact of service user involvement on research: A prospective case study. *International Journal of Consumer Studies, 35*(6), 609-615.

Barber, R., Boote, J. D., Parry, G. D., Cooper, C. L., Yeeles, P., & Cook, S. (2011b). Can the impact of public involvement on research be evaluated? A mixed methods study. *Health Expectations, 15*(3), 229-241.

Barth-Jones, D. C. (2012). The 're-identification' of Governor William Weld's medical information: A critical re-examination of health data identification risks and privacy protections, then and now. *Social Science Research Network*.

Berner, M., Graupner, E., & Maedche, A. (2014). The information panopticon in the big data era. *Journal of Organization Design, 3*(1), 14-19.

Bodnar, N. (2012). U of T News: Predicting the Risk of Death for Heart Failure Patients. Retrieved August 2014, from http://news.utoronto.ca/predicting-risk-death-heart-failure-patients.

BORN (Better Outcomes Registry and Network Ontario). (2014). About BORN. Retrieved August 2014, from https://www.bornontario.ca/en/about-born/.

BORN (Better Outcomes Registry and Network Ontario). (2015). Statement of Information Practices. Retrieved February 2015, from https://www.bornontario.ca/en/privacy/statement-of-information-practices/.

Born, K. & Laupacis, A. (2012). Public engagement in Ontario's hospitals — Opportunities and challenges. *Healthcare Quarterly, 15*(Special Issue), 16-20.

Brook, E. L., Rosman, D. L., & Holman, C. D. A. J. (2008). Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System. *Australian and New Zealand Journal of Public Health, 32*(1), 19-23.

Brownell, M. D., Roos, N. P., Fransoo, R., Roos, L. L., Guèvremont, A., MacWilliam, L.,... Levin, B. (2006). Is the class half-empty? A population-based perspective on socio-economic status and educational outcomes. *IRPP Choices, 12*(5), 1-30.

Brownell, M. D., Chartier, M., Santos, R., Au, W., Roos, N. P., & Girard, D. (2011). Evaluation of a newborn screen for predicting out-of-home placement. *Child Maltreatment, 16*(4), 239-249.

Cabinet Office (The Cabinet Office, Government of the United Kingdom). (2012). *Open Data Strategy*. London, United Kingdom: Government of the United Kingdom.

CaG (CARTaGENE). (n.d.-a). Governance. Retrieved July 2014, from http://cartagene.qc.ca/en/governance.

CaG (CARTaGENE). (n.d.-b). Data Access Procedure. Retrieved November 2014, from http://www.cartagene.qc.ca/en/data-access-procedure.

CaG (CARTaGENE). (n.d.-c). Phase A (Baseline Survey). Retrieved November 2014, from http://www.cartagene.qc.ca/en/phase-baseline-survey.

CAHS (Canadian Academy of Health Sciences). (2009). *Making an Impact — A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*. Ottawa (ON): CAHS.

CAO (Court of Appeal for Ontario). (2012). Jones v. Tsige, 2012 ONCA 32 (CanLII). Toronto (ON): CAO.

Cate, F. H. (2008). *Provincial Canadian Geographic Restrictions on Personal Data in the Public Sector*. Washington (DC): The Centre for Information Policy Leadership, Hunton & Williams LLP.

Caulfield, T., Ries, N., & Barr, G. (2011). Variation in ethics review of multi-site research initatives. *Healthcare, Bioethics and the Law, 3*(1), 85-100.

Cavoukian, A. (2005). *Frequently Asked Questions: Personal Health Information Protection Act*. Toronto (ON): Information & Privacy Commissioner of Ontario.

Cavoukian, A. & El Emam, K. (2011). *Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy*. Toronto (ON): Information and Privacy Commissioner of Ontario.

Cavoukian, A. & Alvarez, R. C. (2012). *Embedding Privacy into the Design of EHRs to Enable Multiple Functionalities — Win/Win*. Montreal (QC): Canada Health Infoway.

Cavoukian, A. & Chanliau, M. (2013). *Privacy and Security by Design: A Convergence of Paradigms*. Toronto (ON): Information & Privacy Commissioner of Ontario.

Cavoukian, A. & Castro, D. (2014). *Big Data and Innovation, Setting the Record Straight: De-identification Does Work*. Toronto (ON): Information & Privacy Commissioner of Ontario.

CBS (Statistics Netherlands). (2004). Guidelines for On Site/Remote Access Output. Retrieved January 2015, from http://www.cbs.nl/en-GB/menu/informatie/beleid/zelf-onderzoeken/dienstencatalogus/default.htm.

CBS (Statistics Netherlands). (2014a). Which Organisations Can Apply for Access to Statistics Netherlands' Datasets for Statistical or Scientific Research? Retrieved September 2014, from http://www.cbs.nl/en-GB/menu/informatie/beleid/zelf-onderzoeken/welke-organisaties-mogen-zelf-onderzoek-doen-op-de-gegevensbestanden-van-het-cbs.htm.

CBS (Statistics Netherlands). (2014b). Microdata Services: Conduct Your Own Research Using Data from Statistics Netherlands. Retrieved September 2014, from http://www.cbs.nl/en-GB/menu/informatie/beleid/zelf-onderzoeken/default.htm?Languageswitch=on.

CCA (Council of Canadian Academies). (2013). *Innovation Impacts: Measurement and Assessment*. Ottawa (ON): The Expert Panel on the Socioeconomic Impacts of Innovation Investments, CCA.

CCO (Cancer Care Ontario). (2012). Synoptic Pathology Reporting. Retrieved August 2014, from https://www.cancercare.on.ca/ocs/clinicalprogs/pathnlabmed/pathproj_prof/.

CCO CPO & CIO (Cancer Care Ontario Chief Privacy Officer and Chief Information Officer). (2011). *Data Use & Disclosure Standard*. Toronto (ON): CCO.

CFHI (Canadian Foundation for Healthcare Improvement). (2014). *Innovative Approach to Patient Engagement Takes Root in Alberta*. Ottawa (ON): CFHI.

Chamberlayne, R., Green, B., Barer, M. L., Hertzman, C., Lawrence, W. J., & Sheps, S. B. (1998). Creating a population-based linked health database: A new resource for health services research. *Canadian Journal of Public Health, 89*(4), 270-273.

Chapman, A. D. (2005). *Principles of Data Quality*. Copenhagen, Denmark: Global Biodiversity Information Facility.

Cheng, T. L., Savageau, J. A., Sattler, A. L., & DeWitt, T. G. (1993). Confidentiality in health care. A survey of knowledge, perceptions, and attitudes among high school students. *Journal of the American Medical Association, 269*(11), 1404-1407.

CHI (Canada Health Infoway). (2012). *Cloud Computing in Health — White Paper*. Toronto (ON): CHI.

CHI (Canada Health Infoway). (2013). *Big Data Analytics in Health: White Paper (Executive Summary)*. Ottawa (ON): CHI.

CHI (Canada Health Infoway). (2014a). *Annual Report 2013-2014*. Toronto (ON): CHI.

CHI (Canada Health Infoway). (2014b). *Connecting Patients with Providers: A Pan-Canadian Study on Remote Patient Monitoring*. Toronto (ON): Ernst & Young LLP.

Christen, P. & Goiser, K. (2007). Quality and complexity measures for data linkage and deduplication. *Studies in Computational Intelligence, 43*, 127-151.

CHSRF (Canadian Health Services Research Foundation). (2011). *Better with Age: Health Systems Planning for the Aging Population: A Backgrounder*. Ottawa (ON): CHSRF.

CIHI (Canadian Institute for Health Information). (2013a). *Adverse Drug Reaction–Related Hospitalizations Among Seniors, 2006 to 2011*. Ottawa (ON): CIHI.

CIHI (Canadian Institute for Health Information). (2013b). *Informing Decisions: Data Improves Rehabilitation Services in Canada*. Ottawa (ON): CIHI.

CIHI (Canadian Institute for Health Information). (2013c). *Health Spending in 2013*. Ottawa (ON): CIHI.

CIHI (Canadian Institute for Health Information). (2014a). Data Quality. Retrieved July 2014, from http://www.cihi.ca/cihi-ext-portal/internet/en/tabbedcontent/standards+and+data+submission/data+quality/cihi021513.

CIHI (Canadian Institute for Health Information). (2014b). *Policy on the Security of Confidential Information and Use of Mobile Devices/Removable Media*. Ottawa (ON): CIHI.

CIHI (Canadian Institute for Health Information). (2014c). Vision and Mandate. Retrieved January 2015, from http://www.cihi.ca/CIHI-ext-portal/internet/EN/SubTheme/about+cihi/vision+and+mandate/cihi010703.

CIHI (Canadian Institute for Health Information). (2014d). Continuing Care Reporting System (CCRS) Metadata. Retrieved October 2014, from http://www.cihi.ca/CIHI-ext-portal/internet/en/document/types+of+care/hospital+care/continuing+care/ccrs_metadata.

CIHI (Canadian Insititute for Health Information). (2014e). ICD-10-CA. Retrieved January 2015, from http://www.cihi.ca/cihi-ext-portal/internet/en/document/standards+and+data+submission/standards/classification+and+coding/codingclass_icd10.

CIHI (Canadian Institute for Health Information). (2014f). Governance and Accountability. Retrieved February 2015, from http://www.cihi.ca/CIHI-ext-portal/internet/EN/SubTheme/about+cihi/governance+and+accountability/cihi010704.

CIHI (Canadian Institute for Health Information). (2014g). Classification and Coding. Retrieved February 2015, from http://www.cihi.ca/CIHI-ext-portal/internet/en/tabbedcontent/standards+and+data+submission/standards/classification+and+coding/cihi010689.

CIHI (Canadian Institute for Health Information). (2014h). *Health Indicators 2013*. Ottawa (ON): CIHI.

CIHR (Canadian Institutes for Health Research). (2005). *CIHR Best Practices for Protecting Privacy in Health Research*. Ottawa (ON): Government of Canada.

CIHR, NSERC, & SSHRC (Canadian Institutes for Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada). (2005). Tri-Council Policy Statement: *Ethical Conduct for Research Involving Humans*. Ottawa (ON): Government of Canada.

CIHR, NSERC, & SSHRC (Canadian Institutes for Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada). (2014). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Ottawa (ON): Government of Canada.

CIHR (Canadian Institutes of Health Research). (2011). *Canada's Strategy for Patient-Oriented Research*. Ottawa (ON): CIHR.

CIHR (Canadian Institutes of Health Research). (2013a). Strategy for Patient-Oriented Research. Retrieved July 2014, from http://www.cihr-irsc.gc.ca/e/41204.html.

CIHR (Canadian Institutes of Health Research). (2013b). Ethics of Health Research Involving First Nations, Inuit and Métis People. Retrieved November 2014, from http://www.cihr-irsc.gc.ca/e/29339.html.

CIHR (Canadian Institutes of Health Research). (2013c). CIHR's Framework for Citizen Engagement - Section Three: The CIHR Citizen Engagement Framework. Retrieved June 2014, from http://www.cihr-irsc.gc.ca/e/41289.html#s3c.

CIHR (Canadian Institutes of Health Research). (2013d). Canadian Longitudonal Study on Aging (CLSA). Retrieved July 2014, from http://www.cihr-irsc.gc.ca/e/18542.html.

Clark, S. & Weale, A. (2011). *Information Governance in Health. An Analysis of the Social Values Involved in Data Linkage Studies*. London, United Kingdom: Nuffield Trust.

CLSA (Canadian Longitudonal Study on Aging). (2009). Home. Retrieved July 2014, from https://www.clsa-elcv.ca/.

CMPA (Canadian Medical Protective Association). (2014). *Electronic Records Handbook*. Ottawa (ON): CMPA.

Colledge, F., Elger, B., & Howard, H. C. (2013). A review of the barriers to sharing in biobanking. *Biopreservation and Biobanking, 11*(6), 339-346.

CPAC (Canadian Partnership Against Cancer). (n.d.). Synoptic Reporting (Surgery). Retrieved January 2015, from http://www.partnershipagainstcancer.ca/priorities/2007-2012-initiatives/cancer-guidelines-2007-2012-strategic-initiatives/synoptic-surgical-reporting-2/.

CPCSSN (Canadian Primary Care Surveillance Network). (2013). About CPCSSN. Retrieved January 2015, from http://cpcssn.ca/about-cpcssn/.

CPTP (Canadian Partnership for Tomorrow Project). (2011). Home. Retrieved July 2014, from http://www.partnershipfortomorrow.ca/cptp/portal/Home?_afrLoop=240101022141000&_afrWindowMode=0&_adf.ctrl-state=52nnybxm4_4.

Curtis, L. H., Brown, J., & Platt, R. (2014). Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Affairs, 33*(7), 1178-1186.

Data Linkage WA (Data Linkage Western Australia). (2014a). *Application and Pricing Arrangements — Data Linkage*. Perth, Western Australia: Government of Western Australia Department of Health.

Data Linkage WA (Data Linkage Western Australia). (2014b). Home. Retrieved August 2014, from http://www.datalinkage-wa.org/.

Davies, C. & Collins, R. (2006). Balancing potential risks and benefits of using confidential data. *British Medical Journal, 333*(7563), 349-351.

Dawson, J. (2006). Privacy and Disclosure of Health Information. In P. D. G. Skegg & R. Paterson (Eds.), *Medical Law in New Zealand*. Wellington, New Zealand: Thomson Brookers.

de Oliveira, C., Nguyen, H. V., Wijeysundera, H. C., Wong, W. W., Woo, G., Grootendorst, P.,... Krahn, M. D. (2013). Estimating the payoffs from cardiovascular disease research in Canada: An economic analysis. *Canadian Medical Association Open Access Journal, 1*(2), E83-E90.

Denham, E. (2013). *Investigation Report F13-02 — Ministry of Health*. Victoria (BC): Office of the Information and Privacy Commissioner for British Columbia.

Denny, J. C. (2012). Chapter 13: Mining electronic health records in the genomics era. *PLoS Computational Biology, 8*(12), 1-15.

DH (UK Department of Health). (1997). *Caldicott Committee: Report on the Review of Patient-Identifiable Information*. London, United Kingdom: Department of Health.

DH (UK Department of Health). (2013). *Caldicott Review: Information Governance in the Health and Care System - Information: To Share Or Not To Share? The Information Governance Review*. London, United Kingdom: Department of Health.

Di Iorio, C. T., Carinci, F., Azzopardi, J., Baglioni, V., Beck, P., Cunningham, S.,... Olympios, G. (2009). Privacy impact assessment in the design of transnational public health information systems: The BIRO project. *Journal of Medical Ethics, 35*(12), 753-761.

Di Iorio, C. T., Carinci, F., Brillante, M., Azzopardi, J., Beck, P., Bratina, N.,... Massi Benedetti, M. (2013). Cross-border flow of health information: Is 'privacy by design' enough? Privacy performance assessment in EUBIROD. *European Journal of Public Health, 23*(2), 247-253.

Dixon-Woods, M. & Tarrant, C. (2009). Why do people cooperate with medical research? Findings from three studies. *Social Science & Medicine, 68*(12), 2215-2222.

Doiron, D., Ferretti, V., Burton, P., Marcon, Y., Gaye, A., Wolffenbuttel, B.,... Minelli, C. (2013a). Data harmonization and federated analysis of population-based studies: The BioSHaRE project. *Emerging Themes in Epidemiology, 10*(12), 1-8.

Doiron, D., Raina, P., & Fortier, I. (2013b). Linking Canadian population health data: maximizing the potential of cohort and administrative data. *Canadian Journal of Public Health, 104*(3), e258-261.

Dormuth, C. R., Hemmelgarn, B. R., Paterson, J. M., James, M. T., Teare, G. F., Raymond, C. B.,... Ernst, P. (2013). Use of high potency statins and rates of admission for acute kidney injury: Multicenter, retrospective observational analysis of administrative databases. *British Medical Journal, 346*, 1-10.

Dormuth, C. R., Filion, K. B., Paterson, J. M., James, M. T., Teare, G. F., Raymond, C. B.,... Lipscombe, L. (2014). Higher potency statins and the risk of new diabetes: Multicentre, observational study of administrative databases. *British Medical Journal, 348*, 1-9.

EEA (European Environment Agency). (n.d.). Environmental Terminology and Discovery Service (ETDS). Retrieved June 2014, from http://glossary.eea.europa.eu/terminology/concept_html?term=acceptable%20risk%20level.

Einav, L. & Levin, J. D. (2013). *The Data Revolution and Economic Analysis*. Paper presented at NBER Innovation Policy and the Economy Conference, Washington (DC).

EKOS (EKOS Research Associates). (2007). *Electronic Health Information and Privacy Survey: What Canadians Think — 2007*. Ottawa (ON): Canada Health Infoway, Health Canada, and the Office of the Privacy Commissioner of Canada.

El Emam, K., Jonker, E., Sams, S., Neri, E., Neisa, A., Gao, T., & Chowdhury, S. (2007). *Pan-Canadian De-Identification Guidelines for Personal Health Information*. Ottawa (ON): CHEO Research Institute.

El Emam, K., Dankar, F. K., Issa, R., Jonker, E., Amyot, D., Cogo, E.,... Bottomley, J. (2009). A globally optimal k-anonymity method for the de-identification of health data. *Journal of the American Medical Informatics Association, 16*(5), 670-682.

El Emam, K., Buckeridge, D., Tamblyn, R., Neisa, A., Jonker, E., & Verma, A. (2011a). The re-identification risk of Canadians from longitudinal demographics. *BMC Medical Informatics and Decision Making, 11*(46), 1-12.

El Emam, K., Jonker, E., Arbuckle, L., & Malin, B. (2011b). A systematic review of re-identification attacks on health data. *PLoS One, 6*(12), 1-12.

El Emam, K., Mercer, J., Moreau, K., Grava-Gubins, I., Buckeridge, D., & Jonker, E. (2011c). Physician privacy concerns when disclosing patient data for public health purposes during a pandemic influenza outbreak. *BMC Public Health, 11*(454), 1-16.

El Emam, K., Samet, S., Hu, J., Peyton, L., Earle, C., Jayaraman, G. C.,... Essex, A. (2012). A protocol for the secure linking of registries for HPV surveillance. *PLoS One, 7*(7), 1-14.

El Emam, K. (2013a). *Guide to the De-identification of Personal Health Information.* Boca Raton (FL): CRC Press.

El Emam, K. (2013b). *Privacy Analytics White Paper: Overview of Re-identification Risk Assessment and Anonymization Process.* Ottawa (ON): Privacy Analytics, Inc.

El Emam, K. (2013c). The return on investment from the de-identification of health data. *Risky Business: The Privacy Analytics Newsletter, March 2013*, 4-7.

El Emam, K. & Arbuckle, L. (2013). *Anonymizing Health Data: Case Studies and Methods to Get You Started.* Sebastopol (CA): O'Reilly Media, Inc.

El Emam, K., Samet, S., Arbuckle, L., Tamblyn, R., Earle, C., & Kantarcioglu, M. (2013). A secure distributed logistic regression protocol for the detection of rare adverse drug events. *Journal of the American Medical Informatics Association, 20*(3), 453-461.

El Emam, K. (2014). The need for anonymization in safe havens. *Risky Business Magazine, October*, 20-24.

El Emam, K. & Malin, B. (2014). *Concepts and Methods for De-Identifying Clinical Trials Data.* Washington (DC): Institute of Medicine Committee on Strategies for Responsible Sharing of Clinical Trial Data.

Etwel, F. A., Rieder, M. J., Bend, J. R., & Koren, G. (2008). A surveillance method for the early identification of idiosyncratic adverse drug reactions. *Drug Safety, 31*(2), 169-180.

EU (European Union). (2000). Charter of Fundamental Rights of the European Union. Retrieved September 2014, from http://www.europarl.europa.eu/charter/pdf/text_en.pdf.

European Commission. (1999). *Opinion of the European Group on Ethics in Science and New Technologies — Ethical Issues of Healthcare in the Information Society, No. 13, 30 July 1999*.

European Commission (2010). Commission Decision of 16 December 2009 laying down guidelines for the management of the Community Rapid Information System 'RAPEX' established under Article 2 and of the notification procedure established under Article 11 of Directive 2001/95/EC (the General Product Safety Directive). *Official Journal of the European Union, 53*, 1-68.

Evans, R. G., McGrail, K. M., Morgan, S. G., Barer, M. L., & Hertzman, C. (2001). Apocalypse no: Population aging and the future of health care systems. *Canadian Journal on Aging, 20*(S1), 160-191.

Fan, J., Lv, J., & Qi, L. (2011). Sparse high dimensional models in economics. *Annual Review of Economics, 3*, 291-317.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review, 1*(2), 293-314.

Farr Institute (Farr Institute of Health Informatics Research). (2015). Innovative Governance. Retrieved February 2015, from http://www.farrinstitute.org/85_Innovative-Governance.html.

FCC (Federal Court of Canada). (2008). *Gordon v. Canada (Health), 2008 FC 258 (CanLII)* Ottawa (ON): FCC.

FDA (Food and Drug Administration). (2010). *The Sentinel Initiative*. Rockville (MD): FDA.

Finnie, R., Childs, S., Pavlic, D., & Jevtovic, N. (2014). *EPRI (Education Policy Research Initiative) Grad Earnings Brief #2: Data and Methodology*. Ottawa (ON): University of Ottawa.

First Nations Centre. (2007). *OCAP: Ownership, Control, Access and Possession. Sanctioned by the First Nations Information Governance Committee, Assembly of First Nations*. Ottawa (ON): National Aboriginal Health Organization.

Flowers, J. & Ferguson, B. (2010). The future of health intelligence: Challenges and opportunities. *Public Health, 124*(5), 274-277.

Forsberg, J. S., Hansson, M. G., & Eriksson, S. (2014). Why participating in (certain) scientific research is a moral duty. *Journal of Medical Ethics, 40*(5), 325-328.

Fortier, I., Burton, P. R., Robson, P. J., Ferretti, V., Little, J., L'Heureux, F.,... Keers, J. C. (2010). Quality, quantity and harmony: The DataSHaPER approach to integrating data across bioclinical studies. *International Journal of Epidemiology, 39*(5), 1383-1393.

Fortier, I., Doiron, D., Little, J., Ferretti, V., L'Heureux, F., Stolk, R. P.,... Burton, P. R. (2011). Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies. *International Journal of Epidemiology, 40*(5), 1314-1328.

Fortier, I., Doiron, D., Wolfson, C., & Raina, P. (2012). Harmonizing data for collaborative research on aging: Why should we foster such an agenda? *Canadian Journal on Aging, 31*(01), 95-99.

Francis, R. (2013a). The Mid Staffordshire NHS Foundation Trust Public Inquiry - Press Statement. Retrieved January 2015, from http://mycouncil.oxfordshire.gov.uk/documents/s20021/HWB_MAR1413R03.pdf.

Francis, R. (2013b). *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. London, United Kingdom: The Stationery Office Limited.

Galbraith, J. R. (2014). Organizational design challenges resulting from big data. *Journal of Organization Design, 3*(1), 2-13.

Gantz, J. F. & Reinsel, D. (2012). *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. Framingham (MA): IDC (International Data Corporation).

Gartner Inc. (2013). Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020. Retrieved August 2014, from http://www.gartner.com/newsroom/id/2636073.

Gaye, A., Marcon, Y., Isaeva, J., LaFlamme, P., Turner, A., Jones, E. M.,... Burton, P. R. (2014). DataSHIELD: taking the analysis to the data, not the data to the analysis. *International Journal of Epidemiology, 43*(6), 1929-1944.

GDQ (Government of Quebec). (2014a). An Act Respecting Access to Documents Held by Public Bodies and the Protection of Personal Information. Retrieved July 2014, from http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=2&file=/A_2_1/A2_1_A.html.

GDQ (Government of Quebec). (2014b). An Act Respecting the Protection of Personal Information in the Private Sector. Retrieved July 2014, from http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=2&file=/P_39_1/P39_1_A.html.

GDQ (Government of Quebec). (2014c). Civil Code of Québec. Retrieved July 2014, from http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=2&file=/CCQ_1991/CCQ1991_A.html.

GenomeCanada & CIHR (Canadian Institutes of Health Research). (2012). Backgrounder: Results of the Genome Canada-CIHR 2012 Large-Scale Applied Research Project Competition in Genomics and Personalized Health. Retrieved September 2014, from http://www.genomecanada.ca/data/Nouvelles/Fichiers%5Cen%5C427_1_Backgrounder%20-%20english.pdf.

Gershon, A. S. & Tu, J. V. (2008). The effect of privacy legislation on observational research. *Canadian Medical Association Journal, 178*(7), 871-873.

Gissler, M. & Haukka, J. (2004). Finnish health and social welfare registers in epidemiological research. *Norsk Epidemiologi, 14*(1), 113-120.

GNB (Government of New Brunswick). (2009). *Personal Health Information Privacy and Access Act, 2009*. Fredericton (NB): Queen's Printer for New Brunswick.

GNB (Government of New Brunswick). (2010). *Personal Health Information Privacy and Access Act General Regulation*. Fredericton (NB): Queen's Printer for New Brunswick.

GNL (Government of Newfoundland and Labrador). (2006). *Privacy Act*. St. John's (NL): Queen's Printer.

GNL (Government of Newfoundland and Labrador). (2013). *Health Research Ethics Authority Act*. St. John's (NL): Queen's Printer.

GNL (Government of Newfoundland and Labrador). (2014). *Personal Health Information Act, 2008*. St. John's (NL): Queen's Printer.

GNS (Government of Nova Scotia). (2010a). Personal Health Information Act, 2010. Retrieved July 2014, from http://nslegislature.ca/legc/bills/61st_2nd/3rd_read/b089.htm.

GNS (Government of Nova Scotia). (2010b). Personal Information International Disclosure Protection Act, 2006. Retrieved July 2014, from http://nslegislature.ca/legc/statutes/persinfo.htm.

GO (Government of Ontario). (2010). Personal Health Information Protection Act, 2004. Retrieved July 2014, from http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_04p03_e.htm.

GO (Government of Ontario). (2012). Freedom of Information and Protection of Privacy Act, 1990. Retrieved July 2014, from http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_90f31_e.htm.

GO (Government of Ontario). (2013). Personal Health Information Protection Act General Regulation, 2004. Retrieved July 2014, from http://www.e-laws.gov.on.ca/html/regs/english/elaws_regs_040329_e.htm.

GOA (Government of Alberta). (2014a). *Health Information Act Designation Regulation, 2001*. Edmonton (AB): Alberta Queen's Printer.

GOA (Government of Alberta). (2014b). *Health Information Act, 2000*. Edmonton (AB): Alberta Queen's Printer.

GOC (Government of Canada). (2014a). *Statistics Act (R.S.C., 1985, C.S-19)*. Ottawa (ON): Minister of Justice.

GOC (Government of Canada). (2014b). *Consolidation - Personal Information Protection and Electronic Documents Act, 2000*. Ottawa (ON): Minister of Justice.

Golle, P. (2006). *Revisiting the Uniqueness of Simple Demographics in the US Population*. Paper presented at Proceedings of the 5th ACM Workshop on Privacy in Electronic Society, Alexandria (VA).

Govt. of BC (Government of British Columbia). (2014a). *Privacy Act*. Victoria (BC): Queen's Printer.

Govt. of BC (Government of British Columbia). (2014b). *Personal Information Protection Act, 2003*. Victoria (BC): Queen's Printer.

Govt. of BC (Government of British Columbia). (2014c). *Freedom of Information and Protection of Privacy Act, 1996*. Victoria (BC): Queen's Printer.

Govt. of MB (Government of Manitoba). (2012). Personal Health Information Regulation, 1997. Retrieved July 2014, from http://web2.gov.mb.ca/laws/regs/current/_pdf-regs.php?reg=245/97.

Govt. of MB (Government of Manitoba). (2014a). The Privacy Act, 2008. Retrieved July 2014, from http://web2.gov.mb.ca/laws/statutes/ccsm/p125e.php.

Govt. of MB (Government of Manitoba). (2014b). The Personal Health Information Act, 1997. Retrieved July 2014, from http://web2.gov.mb.ca/laws/statutes/ccsm/p033-5e.php.

Govt. of SK (Government of Saskatchewan). (1999). *Health Information and Protection Act, 1999*. Regina (SK): The Queen's Printer.

Govt. of SK (Government of Saskatchewan). (2005). *The Privacy Act*. Regina (SK): The Queen's Printer.

GPEI (Government of Prince Edward Island). (2012). *Freedom of Information and Protection of Privacy Act, 1988*. Charlottetown (PE): Queen's Printer.

Gracie, S. K., Lyon, A. W., Kehler, H. L., Pennell, C. E., Dolan, S. M., McNeil, D. A.,... Lye, S. J. (2010). All Our Babies Cohort Study: Recruitment of a cohort to predict women at risk of preterm birth through the examination of gene expression profiles and the environment. *BMC Pregnancy and Childbirth, 10*(1), 1-9.

Graham, D. J., Campen, D., Hui, R., Spence, M., Cheetham, C., Levy, G.,... Ray, W. A. (2005). Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: Nested case-control study. *The Lancet, 365*(9458), 475-481.

Grant, K. (2014, October 16). Hospital Discovers Privacy Breach on Rob Ford's Files, *The Globe and Mail*.

Greenfield, D. L. (2006). Greenberg v. Miami Children's Hospital: Unjust enrichment and the patenting of human genetic material. *Annals of Health Law, 15*(2), 213-249.

Grodin, M. A. & Annas, G. J. (1996). Legacies of Nuremberg: Medical ethics and human rights. *Journal of the American Medical Association, 276*(20), 1682-1683.

Haddow, G., Laurie, G., Cunningham-Burley, S., & Hunter, K. G. (2007). Tackling community concerns about commercialisation and genetic research: A modest interdisciplinary proposal. *Social Science & Medicine, 64*(2), 272-282.

Hadskis, M. (2002). The Regulation of Human Biomedical Research in Canada. In J. Downie, T. Caulfield & C. Flood (Eds.), *Canadian Health Law and Policy* (2nd ed.) Toronto (ON): Butterworths.

Hanlon, P., Lawder, R., Elders, A., Clark, D., Walsh, D., Whyte, B., & Sutton, M. (2007). An analysis of the link between behavioural, biological and social risk factors and subsequent hospital admission in Scotland. *Journal of Public Health, 29*(4), 405-412.

Harris, M. A., Levy, A. R., & Teschke, K. E. (2008). Personal privacy and public health: Potential impacts of privacy legislation on health research in Canada. *Canadian Journal of Public Health, 99*(4), 293-296.

Harris/Decima. (2011). *2011 Canadians and Privacy Survey: Report Presented to the Office of the Privacy Commissioner of Canada.* Ottawa (ON): Office of the Privacy Commissioner of Canada.

HDEC (Health and Disability Ethics Committees). (2012). *Standard Operating Procedures for Health and Disability Ethics Committees.* Wellington, New Zealand: Ministry of Health.

Henke, N., Kelsey, T., & Whatley, H. (2012). Transparency — The Most Powerful Driver of Health Care Improvement? *Health International, Issue 11*, 64-73.

Hertzman, C. P., Meagher, N., & McGrail, K. M. (2012). Privacy by Design at Population Data BC: A case study describing the technical, administrative, and physical controls for privacy-sensitive secondary use of personal information for research in the public interest. *Journal of the American Medical Informatics Association, 20*, 25-28.

Holman, C. D. J., Bass, A. J., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J.,... Stanley, F. J. (2008). A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Australian Health Review, 32*(4), 766-777.

Holmes, D. (2013). Mid Staffordshire scandal highlights NHS cultural crisis. *The Lancet, 381*(9866), 521-522.

House of Lords. (2009). *Genomic Medicine—Volume I: Report.* London, United Kingdom: The Stationary House by Order of the House.

HREA (Health Research Ethics Authority). (n.d.). Home. Retrieved July 2014, from http://www.hrea.ca/home.aspx.

Hui, K.-L. & Png, I. P. L. (2006). The Economics of Privacy. In T. Hendershott (Ed.), *Handbooks in Information Systems, , Vol. 1: Economics and Information Systems.* Bingley, United Kingdom: Emeral Group Publishing Limited.

Hui, K.-L., Teo, H. H., & Lee, S.-Y. T. (2007). The value of privacy assurance: An exploratory field experiment. *Mis Quarterly, 31*(1), 19-33.

ICES & PHO (Insititute for Clinical Evaluative Sciences and Public Health Ontario). (2012a). *Seven More Years: The Impact of Smoking, Alcohol, Diet, Physical Activity and Stress on Health and Life Expectancy in Ontario - Summary*. Toronto (ON): ICES and PHO.

ICES & PHO (Insititute for Clinical Evaluative Sciences and Public Health Ontario). (2012b). *Seven More Years: The Impact of Smoking, Alcohol, Diet, Physical Activity and Stress on Health and Life Expectancy in Ontario*. Toronto (ON): ICES and PHO.

ICES (Institute for Clinical Evaluative Sciences). (2014a). Mission, Vision & Values. Retrieved June 2014, from http://www.ices.on.ca/About-ICES/Mission-vision-and-values.

ICES (Institute for Clinical Evaluative Sciences). (2014b). Frequently Asked Questions. Retrieved January 2015, from http://www.ices.on.ca/About-ICES/FAQs.

ICES (Institute for Clinical Evaluative Services). (2014c). ICES Data & Analytical Services Frequently Asked Questions. Retrieved January 2015, from http://www.ices.on.ca/Data-Services/FAQs.

ICES (Institute for Clinical Evaluative Services). (2014d). Data Access Process. Retrieved January 2015, from http://www.ices.on.ca/Data-Services/Data-Process.

ICES (Institute for Clinical Evaluative Sciences). (2014e). About ICES Research. Retrieved January 2015, from http://www.ices.on.ca/Research/About-ICES-Research.

ICES (Institute for Clinical Evaluative Sciences). (2014f). cd-link. Retrieved February 2015, from http://www.ices.on.ca/Research/Research-programs/Cancer/cd-link.

ICGC (International Cancer Genome Consortium). (2008). *Goals, Structure, Policies & Guidelines*. Retrieved January 2015, from https://icgc.org/icgc/goals-structure-policies-guidelines.

ICGC (International Cancer Genome Consortium). (2012). *Updates to Goals, Structure Policies and Guidelines*. Section E.1 – Informed Consent, Access and Ethical Oversight. Retrieved January 2015, from https://icgc.org/files/ICGC-E1-Apr2013.pdf.

ICGC (International Cancer Genome Consortium). (2014). Home. Retrieved July 2014, from https://icgc.org/.

ICO (Information Commissioner's Office). (2012). *Anonymisation: Managing Data Protection Risk Code of Practice*. Cheshire, United Kingdom: ICO.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124.

IOM. (2000). *To Err Is Human: Building a Safer Health System*. Washington (DC): The National Academies Press.

IOM (Institute of Medicine). (2014). *Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2*. Washington (DC): The National Academies Press.

IPCO (Information & Privacy Commissioner of Ontario). (2010). *Privacy Risk Management - Building Privacy Protection into a Risk Management Framework to Ensure that Privacy risks are Managed, by Default*. Toronto (ON): IPCO.

IPCO & CHEO (Information & Privacy Commissioner of Ontario and Children's Hospital of Eastern Ontario). (2011). *Safeguarding Personal Health Information When Using Mobile Devices for Research Purposes*. Toronto (ON): IPCO.

Ipsos MORI. (2014). *Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes*. London, United Kingdom: Ipsos MORI Social Research Institute.

Ipsos Reid. (2012). *What Canadians Think: Electronic Information and Priacy Survey 2012*. Toronto (ON): Canada Health Infoway.

ISD Scotland. (2010a). eDRIS Frequently Asked Questions — C.1 How Do I Become an Approved Researcher? Retrieved February 2015, from http://www.isdscotland.org/Products-and-Services/eDRIS/FAQ-eDRIS/#c1.

ISD Scotland (Information Services Division, NHS National Services Scotland). (2010b). Becoming an eDRIS User. Retrieved August 2014, from http://www.isdscotland.org/Products-and-Services/EDRIS/Becoming-an-eDRIS-User/.

ISO (International Organization for Standardization). (n.d.). ISO 31000 - Risk Management. Retrieved June 2014, from http://www.iso.org/iso/home/standards/iso31000.htm.

Jenkin, R., Bennett, J., Frommer, M., & Madronio, C. (2006). *A Streamlined National Approach to Scientific and Ethics Review of Multi-Centre Health and Medical Research in Australia*. Paper presented at Australian Health Ministers' Advisory Council (AHMAC) Working Group on a Streamlined National Approach to Ethical and Scientific Review of Multi-centre Research Workshop, Sydney, Australia.

Johansen, H., Bernier, J., Fines, P., Brien, S., Ghali, W., & Wolfson, M. (2009). Variations by health region in treatment and survival after heart attack. *Component of Statistics Canada Catalogue, 20*(2), 29-34.

Joly, Y., Zeps, N., & Knoppers, B. M. (2011). Genomic databases access agreements: Legal validity and possible sanctions. *Human Genetics, 130*(3), 441-449.

Joly, Y., Dove, E. S., Knoppers, B. M., Bobrow, M., & Chalmers, D. (2012). Data sharing in the post-genomic world: The experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO). *PLoS Computational Biology, 8*(7), e1002549.

Jones, J. (2012). *Response Burden: Introductory Overview Lecture*. Paper presented at Fourth International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions, Montréal (QC).

Jones, K. H., McNerney, C. L., & Ford, D. V. (2014). Involving consumers in the work of a data linkage research unit. *International Journal of Consumer Studies, 38*(1), 45-51.

Jutte, D. P., Roos, L. L., & Brownell, M. D. (2011). Administrative record linkage as a tool for public health research. *Annual Review of Public Health, 32*, 91-108.

Kaminska, B. & New, W. (2005). *Wireless Wearable Biomonitors for Lifetime Wellness Optimization*. Paper presented at 3rd IEEE/EMBS Special Topic Conference on Microtechnology in Medicine and Biology, Oahu (HI).

Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2014). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics*, 1-6.

Kho, M. E., Duffett, M., Willison, D. J., Cook, D. J., & Brouwers, M. C. (2009). Written informed consent and selection bias in observational studies using medical records: systematic review. *British Medical Journal 338*, 866-873.

Kirby, B. (2014). *Data Linkage For Pharmacovigilance Using Routine Electronic Health Records*. Paper presented at International Health Data Linkage Conference, Vancouver (BC).

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles (CA): SAGE Publications.

Knoppers, B. M., Abdul-Rahman, M. H., & Bédard, K. (2007). Genomic databases and international collaboration. *King's Law Journal, 18*, 291-312.

Knoppers, B. M., Chisholm, R. L., Kaye, J., Cox, D., Thorogood, A., Burton, P.,... Harris, J. R. (2013). A P3G generic access agreement for population genomic studies. *Nature Biotechnology, 31*(5), 384-385.

Kohane, I. S. (2011). Using electronic health records to drive discovery in disease genomics. *Nature Reviews Genetics, 12*(6), 417-428.

Kosseim, P. & Jospe, D. (2011). *Banking for the Future: "Informing" Consent in the Context of Biobanks*. Paper presented at IV International Seminar on the UNESCO Universal Declaration on Bioethics and Human Rights, Barcelona, Spain.

Kosseim, P., Pullman, D., Perrot-Daley, A., Hodgkinson, K., Street, C., & Rahman, P. (2012). Privacy protection and public goods: Building a genetic database for health research in Newfoundland and Labrador. *Journal of the American Medical Informatics Association, 20*(1), 38-43.

Kosseim, P., Dove, E. S., Baggaley, C., Meslin, E. M., Cate, F. H., Kaye, J.,... Knoppers, B. M. (2014). Building a data sharing model for global genomic research. *Genome Biology, 15*(430), 1-7.

Krist, A. H., Beasley, J. W., Crosson, J. C., Kibbe, D. C., Klinkman, M. S., Lehmann, C. U.,... Waldren, S. E. (2014). Electronic health record functionality needed to better support primary care. *Journal of the American Medical Informatics Association, 21*(5), 764-771.

Kroes, N. (2011). *Opening Up Europe: From Common Standards to Open Data*. Paper presented at OpenForum Europe Summit 2011, Brussels, Belgium.

Kuipers, T. & van der Hoeven, J. (2009). *PARSE Insight: Insight Into Digital Preservation of Research Output in Europe — Survey Report*. Didcot, United Kingdom: PARSE.Insight.

Kuner, C. (2013). *Transborder Data Flows and Data Privacy Law*. Oxford, United Kingdom: Oxford University Press.

Kush, R. & Goldman, M. (2014). Fostering responsible data sharing through standards. *New England Journal of Medicine, 370*(23), 2163-2165.

Lafky, D. (2009). The Safe Harbor Method of De-Identification — An Empirical Test. Washington (DC): Department of Health & Human Services, Office of the National Coordinator for Health Information Technology.

Lapointe, L., Hughes, J., Simkus, R., Lortie, M., Sanche, S., & Law, S. (2012). *The Population Health Management Challenge*. Montréal (QC): St. Mary's Hospital, Medbase Research, and McGill University.

Laurie, G. T. & Sethi, N. (2012). *Information Governance of Use of Health-Related Data in Medical Research in Scotland: Towards a Good Governance Framework*. Edinburgh, United Kingdom: University of Edinburgh.

Laurie, G. T., Jones, K. H., Stevens, L., & Dobbs, C. (2014). *A Review of Evidence Relating to Harm Resulting From Uses of Health and Biomedical Data*. Edinburgh, United Kingdom: Nuffield Council on Bioethics and Wellcome Trust.

Lee, D. S., Stitt, A., Austin, P. C., Stukel, T. A., Schull, M. J., Chong, A.,... Tu, J. V. (2012). Prediction of heart failure mortality in emergent care: A cohort study. *Annals of Internal Medicine, 156*(11), 767-775.

Lemmens, T. & Austin, L. (2009). The End of Individual Control over Health Information: Promoting Fair Information Practices and the Governance of Biobank Research. In J. Kaye & M. Stranger (Eds.), *Governing Biobanks*. Farnham, United Kingdom: Ashgate.

Lemmens, T. (2013). Pharmaceutical knowledge governance: A human rights perspective. *Journal of Law, Medicine & Ethics, 41*(1), 163-184.

Lewis, C., Clotworthy, M., Hilton, S., Magee, C., Robertson, M. J., Stubbins, L. J., & Corfield, J. (2013). Consent for the use of human biological samples for biomedical research: A mixed methods study exploring the UK public's preferences. *BMJ Open, 3*(8), 1-12. doi: 10.1136/bmjopen-2013-003022.

Lewis, S. (2011). How has health services research made a difference? *Healthcare Policy, 6*(Special Issue), 74-79.

Logan, D. (2010, January 11). What is Information Governance? And Why is it So Hard?, *Gartner, Inc.*

Lothen-Kline, C., Howard, D. E., Hamburger, E. K., Worrell, K. D., & Boekeloo, B. O. (2003). Truth and consequences: Ethics, confidentiality, and disclosure in adolescent longitudinal prevention research. *Journal of Adolescent Health, 33*(5), 385-394.

Lyon, D. (2003). Surveillance as Social Sorting: Computer Codes and Mobile Bodies. In D. Lyon (Ed.), *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination.* New York (NY): Routledge.

Maki, K. (2011). Neoliberal deviants and surveillance: Welfare recipients under the watchful eye of Ontario Works. *Surveillance & Society, 9*(1/2), 47-63.

Malin, B. A., Emam, K. E., & O'Keefe, C. M. (2013). Biomedical data privacy: Problems, perspectives, and recent advances. *Journal of the American Medical Informatics Association, 20*(1), 2-6.

Manchester University. (2014). Better Use of Electronic Health Records Makes Clinical Trials Less Expensive. Retrieved July 2014, from www.sciencedaily.com/releases/2014/07/140711091951.htm.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* New York (NY): McKinsey Global Institute.

Marchessault, G. (2011). The Manitoba Centre for Health Policy: A case study. *Health Policy, 6*(Special Issue), 29-43.

Marion, J. & Thomas, B. (2004). *Use of Probabilistic Record Linkage for the Canadian Cancer Registry.* Ottawa (ON): Statistics Canada.

McDonald, A. M. & Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society, 4*, 540-565.

McDonald, S. W., Lyon, A. W., Benzies, K. M., McNeil, D. A., Lye, S. J., Dolan, S. M.,... Tough, S. C. (2013). The All Our Babies pregnancy cohort: Design, methods, and participant characteristics. *BMC Pregnancy and Childbirth, 13*(Supplement 1), S2-S12.

MCHP (Manitoba Centre for Health Policy). (2004). *Diagnostic Imaging Data: The Good, the Bad, and the Potential.* Winnipeg (MB): University of Manitoba.

MCHP (Manitoba Centre for Health Policy). (2010). Journal Publications. Retrieved June 2014, from http://mchp-appserv.cpe.umanitoba.ca/journalPublicationsList.html.

MCHP (Manitoba Centre for Health Policy). (2014a). Advisory Board. Retrieved June 2014, from http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/aboutAdvisoryBoard.html.

MCHP (Manitoba Centre for Health Policy). (2014b). About MCHP. Retrieved January 2015, from http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/about.html.

MCHP (Manitoba Centre for Health Policy). (2014c). Research at MCHP. Retrieved January 2015, from http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/research.html.

MCHP (Manitoba Centre for Health Policy). (2014d). Applying for Access. Retrieved June 2014, from http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/access.html.

MCHP (Manitoba Centre for Health Policy). (2014e). Accreditation 2014. Retrieved January 2015, from http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/projects/media/4_privacy.pdf.

MCHP (Manitoba Centre for Health Policy). (2014f). Population Health Research Data Repository: Available Years of Data. Retrieved December 2014, from http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/protocol/media/Available_Years.pdf.

MCHP (Manitoba Centre for Health Policy). (2014g). Data Quality. Retrieved August 2014, from http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/repository/dataquality.html.

McKenzie, A. & Hanley, B. (2007). *Consumer and Community Participation in Health and Medical Research: A Practical Guide for Health and Medical Research Organisations*. Perth, Australia: The University of Western Australia School of Population Health and the Telethon Institute for Child Health Research.

McNutt, M. (2015). Data, eternal. *Science, 347*(6217), 7. doi: 10.1126/science.aaa5057.

Meagher, N. & McGrail, K. (2013). *Data Access Review Times Study: Report*. Vancouver (BC): PopulationData BC.

Mello, M. M. & Wolf, L. E. (2010). The Havasupai Indian tribe case—Lessons for research involving stored biologic samples. *New England Journal of Medicine, 363*(3), 204-207.

Meslin, E. M. & Cho, M. K. (2010). Research ethics in the era of personalized medicine: Updating science's contract with society. *Public Health Genomics, 13*(6), 378-384.

Milne, R. L., Burwinkel, B., Michailidou, K., Arias-Perez, J. I., Zamora, M. P., Menendez-Rodriguez, P.,... Easton, D. F. (2014). Common non-synonymous SNPs associated with breast cancer susceptibility: Findings from the Breast Cancer Association Consortium. *Human Molecular Genetics, 23*(22), 6096-6111. doi: 10.1093/hmg/ddu311.

Mini-Sentinel. (2014a). Mini-Sentinel: Data Activities. Retrieved December 2014, from http://www.mini-sentinel.org/data_activities/default.aspx.

Mini-Sentinel. (2014b). About Mini-Sentinel: Background. Retrieved December 2014, from http://www.mini-sentinel.org/about_us/default.aspx.

Mosley, M. (2008). *DAMA-DMBOK Functional Framework*. The Data Management Association.

MRC (Medical Research Council). (n.d.). Bona Fide Research. Retrieved August 2014, from http://www.nshd.mrc.ac.uk/data/bona_fide_researchers.aspx.

Murphy, J., Scott, J., Kaufman, D., Geller, G., LeRoy, L., & Hudson, K. (2009). Public perspectives on informed consent for biobanking. *American Journal of Public Health, 99*(12), 2128-2134.

Namjou, B., Keddache, M., Marsolo, K., Wagner, M., Lingren, T., Cobb, B.,... Harley, J. B. (2013). EMR-linked GWAS study: Investigation of variation landscape of loci for body mass index in children. *Frontiers in Genetics, 4*, 1-9.

NCBI (National Center for Biotechnology Information). (2012). GaP FAQ Archive: Applying for Individual-Level Data — Select Datasets by Consent Groups. Retrieved February 2015, from http://www.ncbi.nlm.nih.gov/books/NBK99229/#DArequest.how_to_select_dbgap_datasets_w.

NCBI (National Center for Biotechnology Information). (2013). GaP FAQ Archive: Applying for Individual-Level Data — Step-by-Step Instruction of How to Apply for dbGaP Individual Level Data. Retrieved February 2015, from http://www.ncbi.nlm.nih.gov/books/NBK99229/#DArequest.would_you_give_me_a_stepbystep.

NCBI (National Center for Biotechnology Information). (2015). dbGaP Overview. Retrieved February 2015, from http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html.

New Zealand Ministry of Health. (2014). Mapping Tools — Files to Help with Mapping Between ICD Versions. Retrieved October 2014, from http://www.health.govt.nz/nz-health-statistics/data-references/mapping-tools.

NIH (National Institutes of Health). (2007). Guidance for Developing Data-Sharing Plans for GWAS. Retrieved February 2015, from http://gds.nih.gov/pdf/gwas_data_sharing_plan.pdf.

NIH (National Institutes of Health). (2014). NIH Genomic Data Sharing Policy. Retrieved November 2014, from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html.

Novartis. (2014). Novartis to License Google "Smart Lens" Technology. Retrieved July 2014, from http://www.novartis.com/newsroom/media-releases/en/2014/1824836.shtml.

NPS (National Physician Survey). (2014). *2014 National Physician Survey — Results for Family Physicians*. Mississauga (ON): The College of Family Physicians of Canada, Canadian Medical Association, & The Royal College of Physicians and Surgeons of Canada.

NSS (National Statistical Service). (n.d.-a). Statistical Data Integration Involving Commonwealth Data Glossary. Retrieved June 2014, from http://www.nss.gov.au/nss/home.NSF/pages/Data+Integration+-+Glossary.

NSS (National Statistical Service). (n.d.-b). 2. The Planning Phase - Contents. Retrieved Janaury 2015, from http://www.nss.gov.au/nss/home.nsf/NSS/7A3193EA236B927ACA25763F00096581?opendocument.

NSS PAC (National Services Scotland Privacy Advisory Committee). (2013). Privacy Advisory Committee. Retrieved June 2014, from https://www.wiki.ed.ac.uk/download/attachments/175640383/SHIP+Privacy+Advisory+Committee+Committee+Website+Working+Practices.pdf.

O'Doherty, K. C. & Burgess, M. M. (2008). Engaging the public on biobanks: Outcomes of the BC biobank deliberation. *Public Health Genomics, 12*(4), 203-215.

O'Neill, O. (2003). Some limits of informed consent. *Journal of Medical Ethics, 29*(1), 4-7.

O'Neill, O. (2004). Accountability, trust and informed consent in medical practice and research. *Clinical Medicine, 4*(3), 269-276.

O'Neill, O. (2013). How to Trust Intelligently—TED Blog. Retrieved July 2014, from http://blog.ted.com/2013/09/25/how-to-trust-intelligently/.

OECD (Organisation for Economic Co-operation and Development). (2007). *OECD Recommendation on Cross-border Co-operation in the Enforcement of Laws Protecting Privacy*. Paris, France: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development). (2013a). *Data Protection Principles for the 21st Century - Revising the 1980 OECD Guidelines*. Paris, France: OECD Publishing.

OECD (Organization for Economic Co-Operation and Development). (2013b). *The OECD Privacy Framework*. Paris, France: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development). (2013c). *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*. Paris, France: OECD Publishing.

OECD (The Organisation for Economic Co-operation and Development). (2013d). *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*. Paris, France: OECD Publishing.

Ohno-Machado, L. (2012). To share or not to share: That is not the question. *Science Translational Medicine, 4*(165), 165.

OMA (Ontario Medical Association). (2013). *eHealth Policy Paper — September 2013*. Toronto (ON): OMA.

OPCC, OIPCA, & OIPCBC (Office of the Privacy Commissioner, Office of the Information and Privacy Commissioner of Alberta, Officer of the Information and Privacy Commissioner for British Columbia). (2012). *Getting Accountability Right with a Privacy Management Program*. OPCC, OIPCA, OIPCBC.

OPM & CIPFA (Office for Public Management Ltd. and The Chartered Institute of Public Finance and Accountancy). (2004). *The Good Governance Standard for Public Services - The Independent Commission on Good Governance in Public Services*. London, United Kingdom: OPM and CIPFA.

OSCJ (Ontario Superior Court of Justice). (2001). *Ontario (Attorney General) v. Pascoe, 2001 32755 (ON SCDC)*.

OSCJ (Ontario Superior Court of Justice). (2014). *Evans v. The Bank of Nova Scotia, 2014 ONSC 2135 (CanLII)*. Ottawa (ON): OSCJ.

OSFI (Office of the Superintendent of Financial Institutions Canada). (2010). *Supervisory Framework*. Ottawa (ON): OSFI.

P3G (Public Population Project in Genomics and Society). (2014). Home. Retrieved June 2014, from http://www.p3g.org/.

P3G (Public Population Project in Genomics and Society). (2015). Maelstrom Research Programme. Retrieved January 2015, from http://p3g.org/programmes/maelstrom-research-programme.

Parker-Pope, T. (2008, March 17). Hospital Workers Fired for Snooping on Spears, *New York Times*.

PbD (Privacy by Design). (n.d.). 7 Foundation Principles. Retrieved June 2014, from http://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/.

PBO (Office of the Parliamentary Budget Officer). (2013). *Fiscal Sustainability Report 2013*. Ottawa (ON): PBO.

Peters, S. G. & Buntrock, J. D. (2014). Big data and the electronic health record. *Journal of Ambulatory Care Management, 37*(3), 206-210.

PHRN (Population Health Research Net). (2014). Secure Unified Research Exchange. Retrieved July 2014, from http://www.phrn.org.au/about-us/phrn/sure/.

Ponemon Institute. (2013). *2013 Cost of Data Breach Study: Global Analysis*. Traverse City (MI): Ponemon Institute.

PopData BC (Population Data BC). (2015a). Research in Action: Research Projects. Retrieved February 2015, from https://www.popdata.bc.ca/ria/projects.

PopData BC (Population Data BC). (2015b). The Secure Research Environment. Retrieved January 2015, from https://www.popdata.bc.ca/dataproviders/services/SRE.

PopData BC (Population Data BC). (2015c). Services for Researchers. Retrieved January 2015, from https://www.popdata.bc.ca/researchers.

Pritts, J. (2008). *The Importance and Value of Protecting the Privacy of Health Information: Roles of HIPAA Privacy Rule and the Common Rule in Health Research*. Washington (DC): Institute of Medicine.

Province of B.C. & DataBC. (2014). *Data Custodianship Guidelines for the Government of British Columbia*. Province of British Columbia.

PSC (Public Safety Canada). (2012). *All Hazards Risk Assessment Methodology Guidelines 2012-2013*. Ottawa (ON): PSC.

Purtell, R. A. & Wyatt, K. M. (2011). Measuring something real and useful in consumer involvement in health and social care research. *International Journal of Consumer Studies, 35*(6), 605-608.

Rabin, R., de Charro, F., & Szende, A. (2004). Introduction. In A. Szende & A. Williams (Eds.), *Measuring Self-Reported Population Health: An International Perspective Based on EQ-5D*. Rotterdam, The Netherlands: EuroQol Group.

Raghupathi, W. & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems, 2*(3), 1-10.

RDC (Research Data Centres Program). (2005). *Statistics Canada Research Data Centres (RDCs) - Guide for Researchers Under Agreement with Statistics Canada*. Ottawa (ON): Statistics Canada.

Ring, E. P. (2011). *Report P-2011-002 — October 28, 2011 — Memorial University & Central Regional Health Authority (Baie Verte Miners Registry Project)*. St. John's (NL): Office of the Information and Provacy Commissioner, Newfoundland and Labrador.

Roos, L. L., Menec, V., & Currie, R. J. (2004). Policy analysis in an information-rich environment. *Social Science & Medicine, 58*(11), 2231-2241. doi: 10.1016/j.socscimed.2003.08.008.

Roos, L. L., Brownell, M., Lix, L., Roos, N. P., Walld, R., & MacWilliam, L. (2008). From health research to social research: Privacy, methods, approaches. *Social Science & Medicine, 66*(1), 117-129.

Roos, N. P., Brownell, M., Guèvremont, A., Fransoo, R., Levin, B., MacWilliam, L., & Roos, L. L. (2006). The complete story: A population-based perspective on school performance and educational testing. *Canadian Journal of Education 29*(3), 684-705.

Roos, N. P., Freemantle, J., Farthing, G., & Carr, J. (2011). Taking it to the streets: Figuring out and communicating what's really important in children's health and well-being research. *Healthcare Policy, 6*(Special Issue), 86-87.

Rose, A. F., Schnipper, J. L., Park, E. R., Poon, E. G., Li, Q., & Middleton, B. (2005). Using qualitative studies to improve the usability of an EMR. *Journal of Biomedical Informatics, 38*, 51-60.

Ross, M. K., Wei, W., & Ohno-Machado, L. (2014). "Big data" and the electronic health record. *Yearbook of Medical Informatics*, 97-104.

Rusert, B. (2009). "A Study in Nature": The Tuskegee Experiments and the New South Plantation. *Journal of Medical Humanities, 30*(3), 155-171.

SAIL (Secure Anonymised Information Linkage Databank). (2013). Governance. Retrieved July 2014, from http://www.saildatabank.com/governance.

SAIL (Secure Anonymised Information Linkage Databank). (2014). FAQ. Retrieved July 2014, from http://www.saildatabank.com/faq.

Sankar, P., Mora, S., Merz, J. F., & Jones, N. L. (2003). Patient perspectives of medical confidentiality: A review of the literature. *Journal of General Internal Medicine, 18*(8), 659-669.

SCC (Supreme Court of Canada). (2004). *R. v. Tessling, [2004] 3 S.C.R. 432, 2004 SCC 67*. Ottawa (ON): SCC.

Schick, U. M., McDavid, A., Crane, P. K., Weston, N., Ehrlich, K., Newton, K. M.,... Carlson, C. S. (2013). Confirmation of the reported association of clonal chromosomal mosaicism with an increased risk of incident hematologic cancer. *PLoS One, 8*(3), e59823.

SDU (University of Southern Denmark). (2014). Research: How to Apply for Data. Retrieved January 2015, from http://www.sdu.dk/en/Om_SDU/Institutter_centre/Ist_sundhedstjenesteforsk/Centre/DTR/Researcher.

Seddon, J. J. & Currie, W. L. (2013). Cloud computing and trans-border health data: Unpacking US and EU healthcare regulation and compliance. *Health Policy and Technology, 2*(4), 229-241.

Senate (Standing Senate Committee on Social affairs, Science and Technology). (2014). *Prescription Pharmaceuticals in Canada: Post Approval Monitoring of Safety and Effectiveness*. Ottawa (ON): Senate.

Sethi, N. & Laurie, G. T. (2013). Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. *Medical Law International, 13*(2-3), 169-204.

Sethi, N. (2014). The promotion of data sharing in pharmacoepidemiology. *European Journal of Health Law, 21*, 271-296.

Shey, H. (2013). *Understand The State Of Data Security And Privacy: 2013 To 2014*. Cambridge (MA): Forrester Research.

SHIP (Scottish Informatics Programme). (2012). *A Blueprint for Health Records Research in Scotland*. Edinburgh, United Kingdom: Scottish Government.

SHIP (Scottish Informatics Programme). (n.d.-a). Home. Retrieved July 2015, from http://www.scot-ship.ac.uk/index.html.

SHIP (Scottish Informatics Programme). (n.d.-b). About. Retrieved August 2014, from http://www.scot-ship.ac.uk/about.html.

Sibthorpe, B., Kliewer, E., & Smith, L. (1995). Record linkage in Australian epidemiological research: Health benefits, privacy safeguards and future potential. *Australian Journal of Public Health, 19*(3), 250-256.

Simonite, T. (2013, May 29). Wanted for the Internet of Things: Ant-sized Computers, *MIT Technology Review*.

Sprumont, D., Girardin, S., & Lemmens, T. (2007). The Helsinki Declaration and the Law: An International and Comparative Analysis. In U. Schmidt & A. Frewer (Eds.), *History and Theory of Human Experimentation: The Declaration of Helsinki and Modern Medical Ethics*. Stuttgard, Germany: Franz Steiner Verlag.

SSC (State Services Commission). (2009). *Government Use of Offshore Information and Communication Technologies (ICT) Service Providers: Advice on Risk Management*. Wellington, New Zealand: SSC.

Stanley, F. J. & Meslin, E. M. (2007). Australia needs a better system for health care evaluation. *Medical Journal of Australia, 186*(5), 220-221.

StatCan (Statistics Canada). (2004). Canadian Community Health Survey Cycle 1.2 — Mental Health and Well-Being: Data Dictionary. Master File (Integrated) — Rounded. Retrieved February 2015, from http://www23.statcan.gc.ca/imdb-bmdi/document/5015_D6_T9_V1-eng.pdf.

StatCan (Statistics Canada). (2010a). Microsimulation. Retrieved June 2014, from http://www.statcan.gc.ca/microsimulation/index-eng.htm.

StatCan (Statistics Canada). (2010b). *Projections of the Diversity of the Canadian Population 2006 to 2031*. Ottawa (ON): Government of Canada.

StatCan (Statistics Canada). (2011). Directive on Record Linkage. Retrieved July 2014, from http://www.statcan.gc.ca/record-enregistrement/policy4-1-politique4-1-eng.htm.

StatCan (Statistics Canada). (2012). National Population Health Survey – Household Component – Longitudinal (NPHS). Retrieved February 2015, from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3225.

StatCan (Statistics Canada). (2013). Federal Research Data Centre: Application Process and Guidelines. Retrieved July 2014, from http://www.statcan.gc.ca/rdc-cdr/frdc_application_process_guidelines-eng.htm.

StatCan (Statistics Canada). (2014a). Approved Record Linkages. Retrieved January 2015, from http://www.statcan.gc.ca/eng/record/summ.

StatCan (Statistics Canada). (2014b). How to Access Data. Retrieved January 2015, from http://www.statcan.gc.ca/eng/health/acces.

StatCan (Statistics Canada). (2014c). Canadian Community Health Survey - Annual Component (CCHS). Retrieved February 2015, from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226.

StatCan (Statistics Canada). (2014d). The Research Data Centres (RDC) Program. from http://www.statcan.gc.ca/rdc-cdr/index-eng.htm.

StatCan (Statistics Canada). (2014e). Social Domain Record Linkage Environment - Privacy Impact Assessment. Retrieved February 2015, from http://www.statcan.gc.ca/about-apercu/pia-efrvp/sdle-ecds-eng.htm.

StatCan (Statistics Canada). (2014f). Workshops — Introduction to Public Use Microdata Files (1 Day) 10H0107. Retrieved February 2015, from http://www.statcan.gc.ca/cgi-bin/workshop/wst2.cgi?workshop=40.

StatCan (Statistics Canada). (2014g). Biobank. Retrieved January 2015, from http://www.statcan.gc.ca/eng/survey/household/5071g.

StatCan (Statistics Canada). (2014h). Application Process and Guidelines. Retrieved August 2014, from http://www.statcan.gc.ca/rdc-cdr/process-eng.htm.

StatCan (Statistics Canada). (2014i). Corporate Business Plan. Retrieved January 2015, from http://www.statcan.gc.ca/about-apercu/bp-pe-eng.htm.

StatCan (Statistics Canada). (2015). The Research Data Centres (RDC) Network. Retrieved January 2015, from http://www.statcan.gc.ca/rdc-cdr/network-reseau-eng.htm.

Steinsbekk, K. S., Myskja, B. K., & Solberg, B. (2013). Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem? *European Journal of Human Genetics, 21*(9), 897-902.

Suissa, S., Henry, D., Caetano, P., Dormuth, C. R., Ernst, P., Hemmelgarn, B.,... Teare, G. (2012). CNODES: The Canadian Network for Observational Drug Effect Studies. *Open Medicine, 6*(4), e134-140.

Sweeney, J. (2000). *Simple Demographics Often Identify People Uniquely.* Pittsburgh (PA): H. John Heinz III School of Public Policy and Management, Carnegie Mellon University.

Taylor, L. & Lynch, E. (2010). *Linking Social Care, Housing & Health Data: Data Linkage* Edinburgh, United Kingdom: The Scottish Government.

TBS (Treasury Board of Canada Secretariat). (2010). Guidance on Preparing Information Sharing Agreements Involving Personal Information. Retrieved August 2014, from http://www.tbs-sct.gc.ca/atip-aiprp/isa-eer/isa-eerpr-eng.asp?format=print.

The Scottish Government. (2012). *Joined-Up Data for Better Decisions: Guiding Principles for Data Linkage.* Edinburgh, United Kingdom: The Scottish Government.

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research, 22*(2), 254-268.

Tu, J. V., Willison, D. J., Silver, F. L., Fang, J., Richards, J. A., Laupacis, A., & Kapral, M. K. (2004). Impracticability of informed consent in the registry of the Canadian Stroke Network. *New England Journal of Medicine, 350*, 414-421.

UK Government (Government of the United Kingdom). (2013). *Information: To Share or Not to Share — The Information Governance Review*. London, United Kingdom: Department of Health.

University of Manitoba. (2014). Aging in Manitoba (AIM) Longitudinal Study: Study description. Retrieved January 2015, from http://umanitoba.ca/centres/aging/research/funded_projects/1068.html.

USA PATRIOT Act. (2001). Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism, Act of 2001, Pub. L. No. 107 –56. 115 Stat. 272. Retrieved November 2014, from http://www.gpo.gov/fdsys/pkg/PLAW-107publ56/pdf/PLAW-107publ56.pdf.

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J.,... Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health, 14*, 1144.

Vaughan, G., Pollock, W., Peek, M. J., Knight, M., Ellwood, D., Homer, C. S.,... Sullivan, E. A. (2012). Ethical issues: The multi-centre low-risk ethics/governance review process and AMOSS. *Australian and New Zealand Journal of Obstetrics and Gynaecology, 52*(2), 195-203.

WA Health (Government of Western Australia Department of Health). (2014). *Data Stewardship and Custodianship Policy*. Perth, Australia: WA Health.

Wathieu, L. & Friedman, A. (2007). An Empirical Approach to Understanding Privacy Valuation. *Working Paper (Harvard Business School. Division of Research), 7*(75).

Weber, G. M., Mandl, K. D., & Kohane, I. S. (2014). Finding the missing link for big biomedical data. *Journal of the American Medical Association, 311*(24), 2479-2480.

Weisbaum, K. M., Slaughter, P. M., & Collins, P. K. (2005). A voluntary privacy standard for health services and policy research: legal, ethical and social policy issues in the Canadian context. *Health Law Review, 14*(1), 42-46.

Weise, E. (2014, August 18). Health Information Network Reports 4.5 Million Patients Had Information Hacked, *USA Today*.

WHO (World Health Organization). (2015a). The WHO Family of International Classifications. Retrieved February 2015, from http://www.who.int/classifications/en/.

WHO (World Health Organization). (2015b). International Classification of Diseases (ICD): Revision Steering Group. Retrieved February 2015, from http://www.who.int/classifications/icd/RSG/en/.

WHO (World Health Organization). (2015c). International Classification of Diseases (ICD). Retrieved February 2015, from http://www.who.int/classifications/icd/en/.

Willison, D. J., Emerson, C., Szala-Meneok, K. V., Gibson, E., Schwartz, L., Weisbaum, K. M.,... Coughlin, M. D. (2008). Access to medical records for research purposes: Varying perceptions across research ethics boards. *Journal of Medical Ethics 34*(4), 308-314.

Willison, D. J., Gibson, E., & McGrail, K. (2011). A Roadmap to Research Uses of Electronic Health Information. In C. M. Flood (Ed.), *Data Data Everywhere: Access and Accountability?* Montréal (QC) and Kingston (ON): McGill-Queens University Press.

Wiwchar, D. (2004, December 16). Nuu-Chah-Nulth Blood Returns to West Coast, *Ha-Shilth-Sa*.

Wolfson, M., Wallace, S. E., Masca, N., Rowe, G., Sheehan, N. A., Ferretti, V.,... Little, J. (2010). DataSHIELD: Resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data. *International Journal of Epidemiology, 39*, 1372-1382.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S.,... Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics, 46*(11), 1173-1186.

Yiannakoulias, N. (2011). Understanding identifiability in secondary health data. *Canadian Journal of Public Health, 102*(4), 291-293.

Zika, E., Paci, D., Braun, A., Rijkers-Defrasne, S., Deschenes, M., Fortier, I.,... Ibarreta, D. (2010). A European survey on biobanks: Trends and issues. *Public Health Genomics, 14*(2), 96-103.

Appendices

Appendix A Detailed Overview of Canadian Legal Frameworks

CANADA'S LEGAL FRAMEWORK

Each province in Canada has evolved its own governance frameworks that facilitate access to health data and maintain public trust in and support for research. Health information legislation in all provinces embodies the twin goals of privacy and the promotion of health research. A single health information-specific statute governs all handling of health information in Alberta, Saskatchewan, Manitoba, Ontario, Nova Scotia, New Brunswick, and Newfoundland & Labrador. (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GNL, 2014; GOA, 2014b; Govt. of MB, 2014b). In British Columbia, Prince Edward Island, and Quebec, the regulation of health information is divided between separate "personal information" protection statutes for the public and private sectors (with health information subsumed within "personal" information) (GPEI, 2012; GDQ, 2014a, 2014b; Govt. of BC, 2014c, 2014b). At the federal level, privacy legislation is also relevant in some contexts (GOC, 2014b). However, despite differences in form, certain common elements exist among the provinces (GPEI, 2012; Govt. of BC, 2014b).

Application of Statutes to Custodians of Health Information

Provincial health information legislation typically applies to a wide range of public and private bodies, defined as "custodians," "trustees," or "public bodies" depending on the province. "Health information" is typically defined to encompass all records related to a person's health status, diagnosis history, and treatment history. Of the seven provinces with health information-specific statutes, some use a brief but general definition (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; Govt. of MB, 2014b), while others have broader and lengthier definitions. Among the broadest definitions are found in Alberta and Newfoundland & Labrador. Alberta defines "health information" to include both "diagnostic, treatment and care information" as well as administrative ("registration") information, and the Newfoundland & Labrador definition is nearly as broad in scope (GNL, 2014; GOA, 2014b). In the provinces without health-specific information statutes (British Columbia, Quebec and Prince Edward Island), health information is not defined but considered to be subsumed within "personal information."

Sharing of Health Data that have been De-identified

It is crucial to note that only identifiable health information is regulated by provincial legislation. Custodians and researchers are free to use and share health information that have been de-identified without any legal

constraint. All provincial health information statutes apply only to identifiable health information, and some specifically exempt information that has been "de-identified" from the application of the statute (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; Govt. of MB, 2014b) and permit sharing of it without restriction (GNB, 2009; GOA, 2014b).

How "identifiable" is defined slightly varies across the provinces (see Table 4.1). In some provinces, "identifiable" is used but not defined in the statute (GPEI, 2012; GDQ, 2014a, 2014b; Govt. of BC, 2014c, 2014b). Other provinces use a reasonable-foreseeability test, meaning that information is identifiable only if it is reasonably foreseeable that it could be used, on its own or with other information, to identify a person (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GNL, 2014). Other provinces' definitions of "identifiable" have no reasonable foreseeability component; if the information "allow[s]" identification, it is identifiable (GDQ, 2014a, 2014b; Govt. of MB, 2014b). In Alberta, the test seems the most stringent: information is identifiable only when their identity is "readily ascertainable" from it (GOA, 2014b). Despite these differences, as discussed in Section 4.3, the TCPS specifies a national standard definition for "de-identified" data, which can lend greater consistency in interpretation between Canadian REBs (CIHR *et al.*, 2014). Still, differing visions of "reasonableness" in determining identifiability may exist within even shared definitions.

Table A.1

Provincial Definitions of "Identifiable"

Used but not defined	Reasonable foreseeability test	"Allows" identification	Is identity "readily ascertainable"
• British Columbia • Prince Edward Island • Quebec	• Saskatchewan • Ontario • New Brunswick • Nova Scotia • Newfoundland & Labrador	• Manitoba • Quebec	• Alberta

Duties and Roles of Custodians of Health Information

For identifiable health information, however, custodians of health data are subject to a wide range of statutory duties that are complex and vary somewhat between provinces. Provincial legislation typically prohibits custodians from collecting health information except with the consent of the individual or without consent but authorized by the statute (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GDQ, 2014a, 2014b; GNL, 2014; GOA, 2014b; Govt. of BC, 2014c, 2014b; Govt. of MB, 2014b). Where consent is required for collection, use, or disclosure, legislation typically requires that the consent be informed

and voluntary (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GDQ, 2014b; Govt. of MB, 2014b). In some provinces, a collector of health information must inform the subject individual of the purposes of the proposed collection and use. In most provinces, the legislation provides that consent can also be revoked (Govt. of SK, 1999; GNS, 2010a; GNL, 2014; Govt. of BC, 2014b). If consent is given, most provincial legislation has a general rule that health information must be collected directly from the individual except in prescribed circumstances. One common such exception is where health information is collected from a third party for research purposes (GNS, 2010a; GNL, 2014; GOA, 2014b; Govt. of BC, 2014b, 2014c; Govt. of MB, 2014b). Such provisions are tied to the broader research-facilitation provisions found in provincial legislation, discussed below.

There is also commonality among the provinces in regulating the use of health information collected. Typically, legislation requires the custodian of the health data to only use them for the original stated purposes for which they were collected, or for uses or purposes considered to be consistent with the original purpose (Govt. of SK, 1999; GNB, 2009; Govt. of BC, 2014b, 2014c). As noted in more depth below, uses for research purposes are permitted if the applicable conditions are met (GNS, 2010a; GNL, 2014).

All provincial legislation also imposes duties on custodians to guard health information. Every province imposes obligations on custodians of health information to take steps to ensure it remains confidential (Govt. of SK, 1999; GNB, 2009; GDQ, 2014a, 2014b; GNL, 2014; GOA, 2014b; Govt. of BC, 2014b, 2014c). In some provinces, these duties are worded briefly and generally, typically requiring the taking of necessary measures to ensure security of the information (GPEI, 2012; GDQ, 2014a, 2014b; Govt. of BC, 2014c, 2014b). Other provinces go further and require custodians to develop and comply with written policies, information practices, and safeguards on the holding, sharing, retention, and destruction of health information, as well as designation of a contact person for every custodian (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010). Of all the provinces, though, Alberta's legislation has the most detailed requirements for security controls (GOA, 2014b, 2014a).

In most provinces, custodians must obtain the informed consent of the individual to disclose their health information, but every province's legislation has a wide range of exceptions to this rule, one of which is disclosure for research purposes in accordance with the statute's research-facilitation provisions discussed in the next section (Govt. of SK, 1999; GDQ, 2014a, 2014b; GOA, 2014b; Govt. of BC, 2014c, 2014b; Govt. of MB, 2014b).

In addition to the specific rules for collection, use, and disclosure of health information, most provinces also have overarching duties that require custodians to exercise restraint in collection, use, and disclosure. In most provinces, legislation limits collection, use, and disclosure of health information to what is reasonably necessary for a purpose (Govt. of SK, 1999; GNB, 2009; GO, 2010; Govt. of MB, 2014b). Nova Scotia's legislation limits such collections, uses, or disclosures to the "minimum amount…necessary to achieve the purpose…" (GNS, 2010a), while Alberta's statute refers to an amount of health information "essential" to the purpose (GOA, 2014b). Another common rule is that identifiable health information is only to be used where "other" or de-identified health information will not serve the purpose (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GDQ, 2014b; GNL, 2014; GOA, 2014b).

At the federal level, all of the foregoing elements are found in the *Personal Information Protection and Electronic Documents Act* (PIPEDA), which governs health information collected by federal works, undertakings, or businesses (GOC, 2014b). However, PIPEDA now only applies in four provinces, and likely will be inapplicable in three of those that have private-sector privacy legislation substantially similar to PIPEDA (GOC, 2014b). Until official Orders-in-Council are passed exempting these three provinces from PIPEDA, there could be a possible federal-provincial question about whether PIPEDA or the provincial enactment takes precedence.

Liabilities of Custodians for Breach

In every province, custodians that breach one or more of their statutory duties can be subjected to investigation by the provincial privacy commissioner or other appointed body. In every province, the broad role of the privacy commission or commissioner is to supervise and enforce the privacy statute(s) in the province. Part of that role is to hear complaints against custodians for alleged breaches of privacy or unauthorized collection, use, or disclose. In some provinces, these inquiries can lead to binding orders against the custodian to cease collection, use, or disclosure of health information, or to comply with the statute more generally (GO, 2010; GDQ, 2014b; GNL, 2014; GOA, 2014b; Govt. of BC, 2014b). Other provinces use a lighter approach in which the commissioner makes "recommendations" to the custodian, which the custodian must decide to adopt or reject. In most such provinces, the complaining individual can appeal a custodian's refusal to adopt a recommendation to the courts or another body, which have the power to issue binding orders against the custodian (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GNL, 2014; Govt. of MB, 2014b).

Where health information is wrongly disclosed, custodians can also be liable for damages to the individuals affected, from three basic sources of law. First, the common law: since a 2012 Ontario Court of Appeal decision, it is now widely accepted in Canada that a person can sue for the tort of "intrusion upon seclusion" (CAO, 2012). To succeed, the claimant must prove that the custodian acted intentionally, without lawful justification, and if a "reasonable person would regard the invasion as highly offensive, causing distress, humiliation or anguish" (CAO, 2012). Second, five provinces (British Columbia, Manitoba, Saskatchewan, Quebec, and Newfoundland & Labrador) have statutory provisions creating a general "invasion of privacy" cause of action (Govt. of SK, 2005; GNL, 2006; GDQ, 2014c; Govt. of BC, 2014a; Govt. of MB, 2014a). And third, health information legislation in two provinces (Ontario and British Columbia) provides statutory causes of action against custodians against whom an order has been made by the provincial privacy commissioner (GO, 2010). However, regardless of the legal grounds on which they may be sued, provisions in health information legislation also provide that custodians may avoid liability by showing that they acted in good faith or took all reasonable steps to prevent the breach (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GPEI, 2012; GNL, 2014; GOA, 2014b; Govt. of BC, 2014c; Govt. of MB, 2014b).

In addition to tort liability, custodians in breach of the legislation or an order of the provincial privacy commissioner may also be found to have committed an offence under the legislation (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GDQ, 2014a; GNL, 2014; GOA, 2014b; Govt. of BC, 2014c, 2014b; Govt. of MB, 2014b). In Alberta, researchers also commit an offence if they breach an agreement with a custodian (GOA, 2014b). However, under these provisions, to amount to an offence, a breach by a custodian must have been "wilful." In addition, custodians in three provinces have a defence of "due diligence," or showing that they took all reasonable steps in the circumstances to avoid the breach (GNB, 2009; GNL, 2014; Govt. of MB, 2014b). Thus, if custodians have complied with all of their statutory duties and taken all steps required to secure the agreement of researchers to safeguard the data and carry out the research ethically, they may be shielded from liability in the event of negligence on the part of the researchers.

Research-Facilitating Provisions

Despite all of the foregoing general duties, custodians in every province may disclose identifiable health data to researchers without the consent of the individual if certain requirements are met. Every provincial health information statute has provisions that promote research by permitting the sharing of

identifiable health information without consent when certain conditions are met. While the specific legislative models used vary somewhat from province to province, there are still common features that can be identified.

In all provinces, researchers must obtain approval of their study from a research ethics body of some kind (GNS, 2010a; GO, 2010; GNL, 2013, 2014; GOA, 2014b; Govt. of MB, 2014b). At one end of the spectrum, some provinces directly specify particular bodies to approve health research. In Quebec, for example, research must be approved by the provincial privacy commission (GDQ, 2014a). In Newfoundland & Labrador, it is the Health Research Ethics Board, in Manitoba it is the Health Information Privacy Committee (for government-held records), and in Alberta it is three specific REBs named by regulation (GNL, 2013; GDQ, 2014a, 2014b; GOA, 2014a; Govt. of MB, 2014b). A more flexible approach is found in provinces where the statute allows any REB that has been approved by the minister of health or a government entity to grant research approvals (Govt. of SK, 1999; GNL, 2013, 2014). And in four provinces, the legislation sets standards that REBs must meet to act under the legislation but does not require them to obtain official approval (GNB, 2009, 2010; GNS, 2010a; GO, 2013; Govt. of MB, 2014b).

Table A.2

How Research Ethics Boards are Chosen

Statute designates Privacy Commission(er) as REB	Statute designates specific entities	REB obtains approval of minister or other entity	Standards but no approval requirement	Legislation silent on who may act as REB
• Quebec	• Newfoundland & Labrador • Manitoba • Alberta	• Saskatchewan • Newfoundland & Labrador	• Ontario • New Brunswick • Nova Scotia • Manitoba	• British Columbia

A related point is the variation between provinces in how REBs are regulated and monitored. Again, there is a spectrum of regulation: some provinces designate specific REBs directly by statute (Quebec, British Columbia, Newfoundland & Labrador, Manitoba and Alberta), some provinces regulate REBs with Ministerial or administrative oversight (Saskatchewan, Newfoundland & Labrador), and in others the regulation is left to universities or provincial Colleges of Physicians (Austin & Lemmens, 2009; Lemmens & Austin, 2009).

The administrative steps that researchers must take vary between provinces from simple to complex. In some provinces, the statute does not specify a particular procedure for obtaining research approval from the relevant body

(Govt. of SK, 1999; GNB, 2009; GPEI, 2012; GDQ, 2014a, 2014b; GNL, 2014; Govt. of BC, 2014c, 2014b; Govt. of MB, 2014b). Other provinces, however, have more detailed requirements such as a formal application to the REB, a research proposal, and a research plan including data security measures (GNS, 2010a; GO, 2010, 2013; GOA, 2014b). In Newfoundland & Labrador, an entirely separate statute provides a very detailed and complex approval process: it specifies the steps researchers must take, requires monitoring of research and reporting back to the custodian after the research is complete (GNL, 2013).

On the legal tests and criteria for researchers against which research proposals are assessed, there is some commonality across provinces, although some provinces have more criteria than others (Govt. of SK, 1999; GNB, 2009; GO, 2010; GNL, 2013; GOA, 2014b; Govt. of MB, 2014b). In some provinces, the statute enumerates a lengthy list of legal tests and criteria that must be met, in the opinion of the research review body, before approval can be granted (Govt. of SK, 1999; GNB, 2009; GNS, 2010a; GO, 2010; GOA, 2014b; Govt. of MB, 2014b). In four provinces, a brief provision sets out general criteria for researchers to meet (British Columbia, Quebec, Prince Edward Island, and Newfoundland & Labrador). Taking all of the provinces' legislation into account, researchers are typically required to demonstrate one or more of the following:

- that the benefits of the proposed research outweigh the potential risks to individuals from disclosure of private information (e.g. Govt. of MB (2014b) and Govt. of SK (1999));
- that the researcher is qualified to do the research (e.g. Govt. of BC (2014b) and Govt. of SK (1999));
- the necessity, to the research project, of collecting and using identifiable health information (e.g. GO (2010));
- that they should not be required to obtain consents from the subject individuals because it is "impractical" (e.g. (Govt. of BC, 2014b) and GNS (2010a)), "impracticable" (e.g. GOA (2014b)), "unreasonable, impractical or not feasible" (e.g. Govt. of SK (1999)), "not reasonably practicable" or "unreasonable or impractical", depending on the province (e.g. GNB (2009) and Govt. of MB (2014b));
- that they have adequate measures in place to safeguard the confidentiality of the information collected and used so as to minimize the risks and harm of accidental disclosure (e.g. Govt. of MB (2014b) and GOA (2014b);
- if data matching or data linkage is to take place, that they have ensured that the linkage will not be harmful to the individuals identified and that the benefits from the linkage are in the public interest (e.g. GNB (2009), Govt. of BC (2014c), Govt. of BC (2014b) and GOA (2014b); and
- that the researchers do not have a conflict of interest arising from their funding arrangements or other circumstances.

In addition to meeting these criteria, researchers often must sign an agreement with the custodian before disclosure. There are differences between provinces in whether or not researchers must sign agreements to obtain health data, and in the terms of those agreements. In two provinces, no agreement is required (Newfoundland and Labrador, Quebec), and in others the agreements that researchers must sign range from simple to complex. In some provinces, the agreements are simply to comply with the conditions of disclosure (GO, 2010; GPEI, 2012; Govt. of BC, 2014c, 2014b). In three provinces, the agreements extend to non-publication, adherence to purpose, and measures to secure data (Govt. of SK, 1999; GNB, 2009; Govt. of MB, 2014b). The most complex agreements researchers must sign are in Nova Scotia and Alberta, where the legislation sets out an extensive list of duties as terms of the researcher agreement (GNS, 2010a; GOA, 2014b). In Alberta, the agreement is also enforceable by Court order. Once all of these steps have been completed, custodians are then authorized to disclose the health information sought by the researchers.

Table A.3

Researcher-Custodian Agreement Requirements

No agreement required	Simple and general in terms	Statute requires detailed terms	Statute prescribes lengthy, detailed list of terms
• Newfoundland & Labrador • Quebec	• British Columbia • Prince Edward Island • Ontario	• Saskatchewan • Manitoba • New Brunswick	• Alberta • Nova Scotia

Aside from the responsibilities that researchers take on under the agreements they sign with custodians, the statutory duties of researchers in receipt of health data range from few to many depending on the province. On one end of the spectrum, in most provinces, researchers are not bound by the same general statutory duties as other custodians or trustees, and in Newfoundland & Labrador, the legislation specifically states that disclosure to a researcher does not make the researcher a "custodian" under the statute (GNL, 2014). This frees researchers from compliance with the wide range of confidentiality and other duties applicable to custodians or trustees. At the other end of the spectrum, in New Brunswick and Saskatchewan, researchers in receipt of health information are subjected to the same duties as custodians (Govt. of SK, 1999; GNB, 2009; Govt. of MB, 2014b). In Ontario and Newfoundland & Labrador, legislation imposes specific duties on researchers in receipt of health data (GO, 2010; GNL, 2013).

Role of Designated Research Entities

In some provinces, custodians may disclose to specially designated entities. In Ontario, the legislation permits disclosure of health information to a "prescribed entity" (GO, 2010, 2013), which may use and disclose health information for research purposes (GO, 2013). Currently, these entities are Cancer Care Ontario, the Canadian Institute for Health Information, the Institute for Clinical Evaluative Sciences, and the Pediatric Oncology Group of Ontario. In Manitoba, regulations designate the Manitoba Centre for Health Policy at the University of Manitoba and the Canadian Institute for Health Information as "prescribed health research organizations" that are able to collect health information for research related to a series of purposes (Govt. of MB, 2012, 2014b). In Alberta, the legislation provides that health information may be disclosed without consent to a designated "health information repository" (GOA, 2014b) however, no such entity has yet been designated by regulation.

Data Linking Rules

Whether done by custodians or researchers, data linking can be an important research strategy but carries risks of identification of a person by combining information together. In four provinces, the legislation specifically regulates "data linking" or "data matching" with health information (GNB, 2009; GNS, 2010a; GOA, 2014b; Govt. of BC, 2014c). In these provinces, it is usually defined as the creation of identifiable health information from the combination of two or more other pieces of information, whether de-identified or not. In Alberta and Nova Scotia, researchers must submit an explanation or assessment of the need for data matching (GNS, 2010a; GOA, 2014b). In Alberta and New Brunswick, custodians may only do data linking with information that arose from approved research (GNB, 2009; GOA, 2014b). In British Columbia, regulations governing data linking by public bodies have not been enacted yet (Govt. of BC, 2014c).

Prospects for Interprovincial Health Data Sharing

Can researchers collect and pool data from custodians in multiple provinces in pursuit of national-level studies? If the data are all de-identified, then yes. If not, then whether pooling can take place depends on the provisions of provincial legislation that apply to this issue.

Again, there is a range of provincial regulation on out-of-province disclosures. At one end of the spectrum, some provinces' legislation is silent on out-of-province disclosures of health information (Govt. of SK, 1999; GO, 2010; GPEI, 2012; Govt. of MB, 2014b). In four other provinces, out-of-province sharing is *prima facie* prohibited but permitted if authorized by the statute's

research-facilitation provisions (GNB, 2009, 2010; GNS, 2010a; GNL, 2014; Govt. of BC, 2014c). And the most restrictive are in Quebec and Alberta, where the statutes prohibit out-of-province disclosures unless the custodian ensures that the receiving province's laws provide equivalent privacy protections for the health information (Quebec) or "takes reasonable steps to protect" the information to be shared (GDQ, 2014a, 2014b; GOA, 2014b).

Table A.4

Provincial Regulation of Interprovincial Data Sharing

No restrictions	Permitted for research purposes	Requirement to ensure protection of data
• Ontario • Saskatchewan • Manitoba • Prince Edward Island	• British Columbia • Nova Scotia • Newfoundland & Labrador • New Brunswick	• Quebec • Alberta

However, for such sharing to occur, the researchers must meet the requirements of their home province, and the donors must be authorized in their province to disclose for research purposes. This necessitates seeking approvals in multiple jurisdictions for the same research study. As noted above, there are interprovincial variations in the criteria that researchers must meet for approval. While some standards are common between all provinces, some provinces do not have the same range of requirements as others. In addition, even when two provinces apply identical criteria for research approval, there is still a risk of inconsistent interpretations between those provinces on key criteria such as "reasonable necessity" for identifiable data. Thus, research approved in a destination province may not meet the legislative standards in the province from which disclosure is sought.

However, this problem could be overcome if researchers have the approval from the REB In their home province recognized as a valid approval in other provinces. In some provinces, this appears possible. In Ontario, approval from an out-of-province REB is recognized and qualifies to permit disclosure in the same way as an Ontario REB (GO, 2010). In other provinces, the definition and requirements of REBs are worded broadly enough to encompass out-of-province entities (GNB, 2009, 2010; GNS, 2010a). And in those provinces where REBs must be approved by government or a government agency, it remains theoretically possible on the language of the statute to qualify an out-of-province REB for such approval. Therefore, in most provinces it is possible to collect health information under the authority of an REB approval granted in another province. By contrast, in Alberta, Quebec, and Manitoba's public sector, REBs or other research approval entities are prescribed by statute, so out-of-province researchers will require their approval. A related problem, noted above, is

that REBs in some provinces are more tightly regulated than in others, which could lead REBs to be skeptical of the approval of out-of-province REBs that they consider to be less well regulated.

Prospects and Challenges for International Data Sharing

Similarly, there are potential legal obstacles to the increasing occurrence of data sharing between countries and across international borders. Technological advances and investments in infrastructure and international consortia are fostering the creation of global research networks such as disease cohort studies, population health data repositories, and biobanks, all of which engage in data sharing.

In particular, the rapid growth of cloud-based computing in health research has, in a sense, globalized the sharing of data. Cloud-based computing has come to mean storage of data on remote servers potentially accessible worldwide by authorized users. Though cloud-based data sharing has obvious advantages for researchers, concerns have arisen about unauthorized users having access to data and unauthorized reuses for which consent has not been obtained from patients or research participants. A related concern is how data stewardship responsibilities will be carried out in a cloud environment that may implicate a number of people and entities (e.g., what happens in the event of a data breach or what happens to data at end-of-life of a project). The global nature of the cloud also means that it is difficult to know which laws apply, let alone how to ensure compliance with the applicable laws.

In addition to challenges posed by new technologies, more funding agencies (in addition to other entities such as journals) are imposing data sharing requirements on researchers, including the deposit of data in repositories that are either entirely open to the public or are available to a limited number of qualified researchers (Ohno-Machado, 2012). Together, the effect of this changing landscape is that data are swiftly moving away from the laboratory or office in which they were generated and into devices and databases around the world.

International and domestic privacy laws are therefore playing a bigger role in international data sharing. At the international level, guidance includes the OECD *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (OECD, 2013c), and at the national level are laws such as Canada's federal and provincial privacy legislation. The potential issues arising from sharing data internationally, especially personal health information, include:

- non-compliance with national/provincial laws, particularly privacy laws;
- unauthorized release of personal information;
- inability to provide individuals with access to their personal information;
- inability to cooperate with national/provincial regulators regarding complaints;

- inability of the national/provincial regulators to investigate or enforce laws;
- inability to guarantee the protection of personal information in countries without privacy or data protection laws;
- conflicts between foreign laws and national/provincial laws;
- possible access to sensitive data by foreign governments;
- overseas judicial decisions that might require the disclosure of sensitive data;
- problems with recovery or secure disposal of data; and
- loss of trust if data are transferred and misused.

(SSC, 2009)

Each of these issues can unduly impede the flow of health and social data and harm valuable health research and health system innovation. A recent global review of the literature on biobanks found that legal issues were the most frequently cited barriers to data sharing (Colledge *et al.*, 2013). Other studies on perceived barriers to data sharing routinely find legal issues rank as the top problem (Kuipers & van der Hoeven, 2009; Zika *et al.*, 2010).

An international comparison of privacy legislation showed that different countries or regions follow different approaches to privacy and data protection. In Europe, for example, data protection is treated as a "fundamental right" whereas in other jurisdictions, such as the United States, data protection is not considered a fundamental human right. Additionally, depending on the jurisdiction, international data sharing regulation can be based on the level of data protection in the country to which the data are imported (e.g., European countries), the measures taken by the person or organization that receives the data (e.g., Canada), or the ability to realize the benefits of electronic commerce (Asia-Pacific Economic Cooperation region) (Kuner, 2013).

This disharmony in legislation between countries carries repercussions for researchers. Researchers within and outside Canada may have to navigate multiple laws and regulations with different standards for access to and use of data. One research study found, for instance, that current provincial legislative restrictions on international data sharing causes "fewer services available to Canadian public bodies […], increased bureaucracy and significantly reduced efficiency, higher financial costs, the threat of tangible harms to health and safety, and the undermining of competition for public bodies' business and of Canada's burgeoning services industry" (Cate, 2008).

When data are shared internationally via an agreement with a non-Canadian entity, a recurring issue is determining the scope of jurisdiction over data breaches outside Canada. Once personal information is in the custody and control of an entity in a foreign jurisdiction, the laws of Canada are unlikely to apply, though

it is possible that data custodians may have a continuing legal responsibility for the confidentiality of health information disclosed to, or stored, or used by, a person in a foreign jurisdiction. However, there has been improvement in cooperation between regulators outside of traditional legal assistance channels, such as through APEC's Cross-border Privacy Arrangement (APEC, 2014), which creates a framework for regional cooperation in the enforcement of privacy laws (and of which the Office of the Privacy Commissioner of Canada is a participant), and through international dispute resolution mechanisms that allow individuals and organizations to assert their privacy rights overseas.

Another issue arising from international data sharing is the adequacy of consents given by research subjects. All privacy laws in Canada permit the international sharing of data if the research subject provides consent. Yet obtaining express consent in writing for international data sharing for health research purposes can be a challenging endeavour. Consent forms provided to research participants (to say nothing of patients in a clinical care context) may not be "future proofed" for all possibilities, including sharing of their (personal) information for other research projects, perhaps with a purpose distinct from the original collection. Part of this is attributable to research projects carried out at a time before wide-scale data sharing was envisaged. This leaves the legality of internationally sharing data from these research projects in doubt.

Besides consent issues, international data sharing also encounters the problem of a multiplicity of institutional data sharing/access policies, Researchers often encounter different access procedures each time they apply for access to data from different institutions, and access agreements often are drafted in highly legalistic terms that fail to communicate expected commitments clearly and understandably to researchers. Some international research consortia are seeking to overcome such problems by harmonizing data sharing agreements. For example, the Public Population Project in Genomics and Society (P3G) (P3G, 2014) has developed a Generic Access Agreement to improve transparency and interoperability in the international sharing of data for use by population genomic studies (Knoppers *et al.*, 2013).

Roles of Governance Frameworks

The provincial legislative frameworks discussed above are sometimes very specific and in other instances leave open key concepts and terms for definition by decision-makers. In the latter instances, governance frameworks can play a useful role in developing harmonized national standards and definitions so as to facilitate interprovincial cooperation among researchers. One key concept is that of "de-identified" data. As noted, such data may legally be shared and used by researchers without restriction, but the legislative definition for

de-identified data varies between provinces: some have no definition, others use a reasonable-foreseeability test, and in Alberta a "readily ascertainable" test is used. Researchers and custodians will benefit from a consistent national approach to what kinds of information should be removed from health data to make them de-identifiable under the law. As noted above, the TCPS has stipulated a national shared definition of "identifiable" data for use by REBs, so this may increase consistency in research review between provinces (CIHR *et al.*, 2014).

Another concept requiring a shared meaning is the necessity for using identifiable health information in research. REBs in different provinces may come to different conclusions on this issue for the same research, with one REB finding that de-identified data will suffice and another REB that it will not. A related term common to the use of health information in research is whether it is "impracticable" to obtain consents from the subjects. Only Nova Scotia specifically defines this term (to mean "a degree of difficulty higher than inconvenience or impracticality but lower than impossibility" (GNS, 2010a). If REBs in different provinces adopt different views of this term, this may create a barrier to interprovincial data sharing.

Assessment of Canadian Provincial Legislative Models

Each province's legislation strives in different ways to protect health information and promote health research. In some provinces, there are more privacy protections than research-facilitation provisions, tipping the balance in favour of privacy and a possibly more restrictive view of the kinds of research that are appropriate. In others, the research-facilitation provisions are extensive but lack enough checks and balances to ensure that privacy values are fully respected. In the Panel's view, the best practices involve the maintenance of strong privacy rules in tandem with rules giving an incentive to health researchers without unduly interfering with privacy values.

Another key best practice involves the facilitation of interprovincial data sharing with the Newfoundland & Labrador framework the best example. There, the legislation provides the same level of protection for health information as other provinces, but has superior provisions for promoting public trust in health research. It has clear mechanisms and procedures researchers must use to seek approval, as well as flexibility in the manner in which research is approved. Researchers may seek approval either from the standing provincial Health Research Ethics Board or a different REB, possibly one out of province that obtains approval from the provincial Health Research Ethics Authority. At the same time, researchers must be overseen by the approving REB and must report back to it when the project is complete. If other provinces followed this model, the prospects for nationally based studies would improve.

Appendix B High-Level Description of an Effective De-Identification Process

De-identification will result in some loss of accuracy or usefulness of the data, but the degree of such loss can be adjusted depending on who uses data in what circumstances. If, for example, those receiving the data have a track record of using data for research, sign confidentiality contracts, and hold data in secure locations, then the degree of de-identification can be less. A well-documented risk-based approach to de-identification is the cornerstone of enabling access to individual-level data with minimal risk of re-identification while maintaining data utility.

A key test for determining whether de-identification has been sufficient is what information a "motivated intruder" could garner from the data set that can be accessed. The threshold for what this information might be is set at a high level. Many attacks that have been made on a database have raised attention because a single record has been identified correctly (Barth-Jones, 2012). In such instances, there is no monetary reward for a successful attack, but such successes undermine public trust that confidentiality is being respected. A key metric, consequently, is to lower the probability that the data to which access is granted would allow a single record to be associated with an individual.[21] To heighten this threshold, the attacker is assumed to know that a person is in the database, which may be obvious in a province-wide database but not for more localized databases. Another strategy to control against is that the attacker has the data and wishes to identify a person from the data in the real world. These risks may pertain to an employee of the data custodian who recognizes a neighbour, which is why the number of individuals who come into contact with individual data needs to be limited.

Once the content of each record is scrutinized for data fields that could include identifiable information, the process of de-identification removes or disguises these fields. Those fields that are thus adjusted mean that there is a reduced chance of associating the released record with other information that is in the public domain or otherwise known by an adversary. Hence, for example, if the accessible health records included an address of Sussex Drive, Ottawa and the person was born in April 1959, then such information could be associated with

21 Looking at the risk of re-identifying a single record is a high threshold since those with criminal intent would need to obtain many records in order to even hope of garnering a financial return for the cost of attempting to undertake a re-identification attack.

other information in the public domain to find out who the person was and then their health data would be identifiable. The de-identification process removes or perturbs such identifiers.

After data custodians de-identify databases appropriately, the released data record contains less detailed information so, for example, a record to be released may show the patient only as a female born in Ottawa in 1941. Critically, there are many other people in the database who also have those characteristics, so identifying the record with a particular person is difficult. Hence, the number of individuals in the released database with similar characteristics is a central concept in determining how much de-identification can be done. If the attacker knew that Anne Dupont was born in Ottawa in 1941 and knew that she was in the database released, but there are four other people in the released data that are also female and born in Ottawa in 1941, then the probability that a single file is that of Anne Dupont is one in five, which is likely of limited use to the attacker.

This probability is known as the *re-identification risk*, and is a key measure that can be set based on the risk involved in releasing the data. In the example above, releasing data to a researcher with birth years adjusted to indicate the person was born between 1940 and 1950 and was a resident of Ontario would dramatically increase the number of people with the same characteristics in the data, and hence reduce the re-identification risk.

Sweeney (2000) showed from an analysis of census data that 87% of the U.S. population is unique — which makes re-identification easier — through the combination of gender, date of birth, and zip code. However, even limited de-identification could significantly reduce the possibility of identifying individuals. If census data contained only the month and year of birth, the percentage of those uniquely identifiable drops to 4.2%. Further replacement of the zip code with the county of residence reduces the population that is uniquely identifiable to 0.2% (Golle, 2006).

For Canada, El Emam *et al.* (2011a) estimated that 98% of the population of Montréal is unique based on full postal code, date of birth, and gender. If the postal code is limited to three characters and combined with the full date of birth, 80% of the population is unique. These results indicate that care needs to be exercised when enabling access to data. Given these known risks, if a database that included information on full postal code, gender, and birth date were shared, a successful re-identification attack would not be surprising.

To estimate the probability of a successful re-identification attack, the U.S. Department of Health and Human Services conducted a study of data that has undergone de-identification following U.S. standards. Using a data set that included about 15,000 records, an attack was simulated by manually searching through external data to see if any of the records aligned to identify a unique person. Two individuals were identified successfully, for a success rate of less than 0.013%. In weighing the cost of such an effort against the value of success based on this probability, there is little incentive to attempt to re-identify individuals in data sets that have undergone rigorous de-identification (Lafky, 2009).

In practice, databases include many fields that could be used to associate files with other information that is in the public domain or otherwise known to an adversary. Scrutinizing the data for such identifiers involves specialized knowledge of whether these fields are replicable, distinguishable, and knowable, e.g., whether the information is known to an acquaintance or available in a public database such as a voter registry. In turn, the data from these fields can be aggregated, masked, or coded, or have some of their values suppressed to lessen re-identification risk.

Re-identification risk can also be adjusted based on a wide range of factors such as who is to receive the data, how they will handle the data, what incentives they have to re-identify someone from the data, and whether there are contracts in place to ensure the data recipient is liable for poor data handling. If sensitive data (such as the records of individuals with HIV/AIDS infection) are to be released as a public database for online access, the data could be de-identified such that re-identification risk was extremely low. For less sensitive data to be held in a secure facility, the re-identification risk could be set at a higher level (El Emam, 2013b).

Many well-publicized successful re-identification "attacks" involved data sets that do not meet the latest standards for de-identification; the data sets released contained fairly precise identifying characteristics than could be matched with other publicly available data, for instance El Emam *et al.* (2011b) and Barth-Jones (2012).

Cavoukian and Castro (2014) conclude that "the risk of re-identification of individuals from properly de-identified data is significantly lower than indicated by commentators on the primary literature." As noted for the United Kingdom, "[t]here is a perception that too much information is being disclosed inadvertently as well as too little being shared deliberately" (DH, 2013).

Given that there is a risk — albeit low — to privacy from release of data that have undergone de-identification, what would be the attendant harm if that risk materialized? Placing value on the harm of breaching privacy in general is difficult. Although polling and the media have highlighted public concerns over a loss of privacy (e.g., Harris/Decima (2011)), researchers monitoring the actual behaviour of individuals have found little evidence that the public believes there are large costs of lost privacy (Hui & Png, 2006; Tsai *et al.*, 2011). The researchers would have expected greater behaviour change if the costs of lost privacy were perceived as large. Behavioral changes to protect privacy, have, however, been documented (see Section 3.4.2).

Exploring this dichotomy has led researchers to uncover complex ways in which people value privacy and in which people differ in their valuation of privacy. Valuations differ according to precise situations and the use to which information is to be put. For example, individuals are more ready to agree to divulge personal information if accessing a website is important to them. In an experiment in Singapore, Hui *et al.* (2007) found that people were willing to disclose more personal information in exchange for small monetary incentives.[22] The type of information that is revealed can affect the valuation, as can the circumstances in which it is revealed. Research based on experiments reported in Wathieu and Friedman (2007) suggested that "consumers are sensitive to context and indirect effects, rather than data collection itself." The authors suggest that consumers' concerns do not centre on pieces of information but focus instead on the use to which the data are put. Acquisti *et al.* (2013) concluded that "what people decide their data is worth depends critically on the context in which they are asked, and specifically on how the problem is framed." For an example of tangible behavioural changes caused by privacy concerns about health information use, see Section 3.4.2.

A recent report examining evidence of harms arising from data linkage and prepared for the Nuffield (UK) Council on Bioethics notes a distinction between legally recognized harms for which remedies might be available, on the one hand, and softer psychological impacts of inappropriate uses of personal data, for which no legal recourse is available, and yet which might have considerable psychological consequences, on the other. The report concludes:

> Given that the abuse of data can result in multiple types of harm (financial, legal, physical, social, and psychological), the prevention of harmful processing and/or award of damages can only address a small aspect of harm caused to individuals. Furthermore, these remedies cannot

22 Similar findings are reviewed in Acquisti (2010) and Acquisti *et al.* (2013).

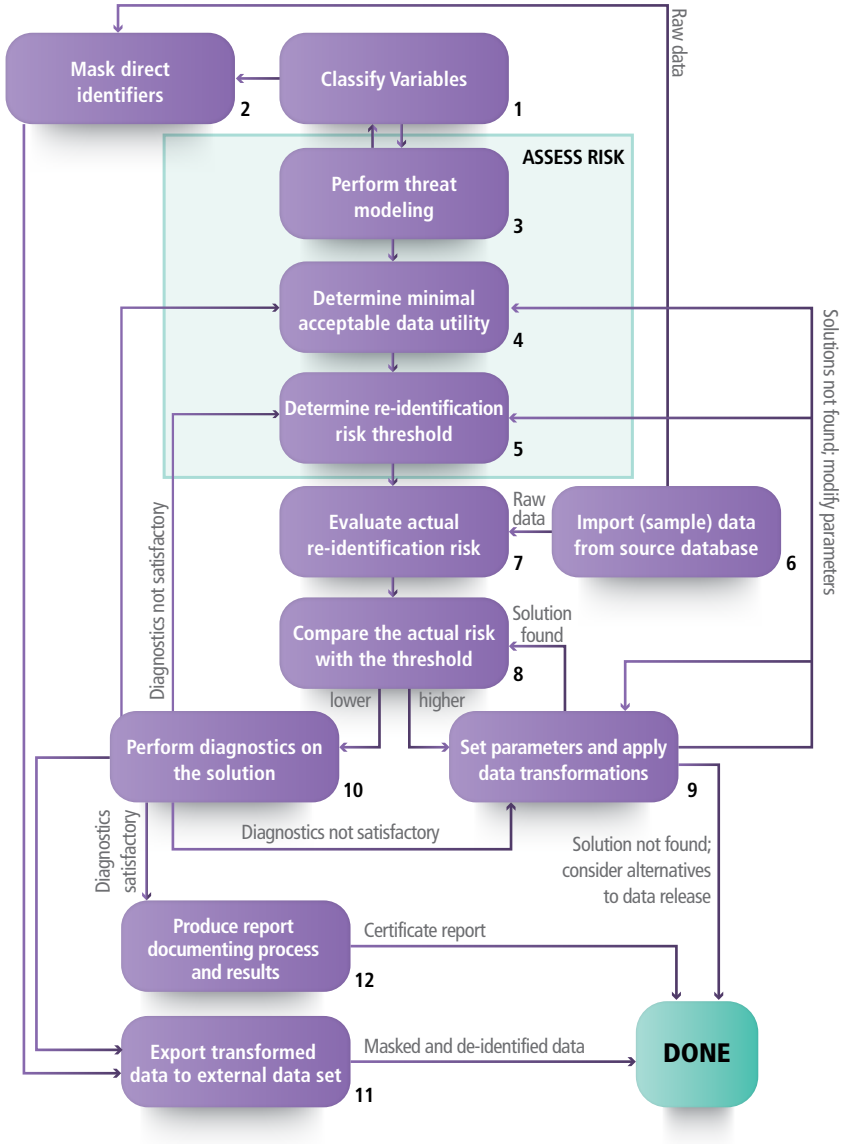
> rectify harm caused to broader public interests such as diminishment of public trust in the health services they receive or in the confidentiality of relationships e.g. between doctors and patients.
>
> (Laurie *et al.*, 2014)

The implications of this are that "[m]ere compliance to legal rules or official guidance might not be enough to secure the social license required for trusted and effective data use, linkage, sharing and transfer. A governance system that shows awareness of, and responsiveness to, likely impacts of data management is more likely to meet this objective" (Laurie *et al.*, 2014).

A further harm from data being used inappropriately is to the data custodians themselves, which may face legal liability and damage to their reputation. The Ponemon Institute estimated the cost of a data breach to a private company by surveying firms on their resulting legal costs, costs to business processes, etc. Across nine countries, the average cost was USD$136 per record, reaching USD$199 per record in Germany (Ponemon Institute, 2013). Strategies to reduce these risks are identified and covered in Chapter 5.

The picture that emerges from analysis of actual individual behaviour is highly nuanced and stands in contrast to positions that privacy is to be protected at all costs. Although most of the analysis has tried to evaluate the value of privacy in general rather than privacy specifically of health and health-related data, the same principles would apply. In particular, individuals appear to be open to data being used if they believe the use is appropriate and justified. For example, well over 90% of respondents to Statistics Canada's Canadian Community Health Survey agree to allow Statistics Canada to link and share their personal, individual-level survey data (StatCan, 2004).

The process model in Figure B.1 illustrates the complete steps for de-identifying a data set in practice. The assumption of that process is that a data set will be shared with a Qualified Investigator, who in this case would be a health researcher.



Reproduced with permission from El Emam (2013b)

Figure B.1

The Overall De-Identification Process for Re-Identification Risk Assessment and De-Identification

This figure shows the steps involved in de-identifying a data set. It illustrates a general process that is intended to be suitable under different use cases. The process would be followed each time a data set needs to be de-identified. Each step is described in detail below.

"Step 1: Classify variables

Determine which fields in the data set are direct identifiers, quasi-identifiers, and not of interest from a de-identification perspective.

Direct identifiers refer to attributes which can uniquely identify an individual, either by themselves or in combination with other readily available information. For example, there are more than 200 people named "John Smith" in Ontario (based on a search in the public telephone directory), therefore the name by itself would not be directly identifying, but in combination with the address it would be directly identifying information. A telephone number is not directly identifying by itself, but in combination with the readily available public telephone directory it becomes so. These numbers are identifying because there exist public and/or private databases that an adversary can plausibly get access to where these numbers can lead directly, and uniquely, to an identity.

The quasi-identifiers are the background knowledge variables about individuals in the disclosed data set that an adversary can use, individually or in combination, to probabilistically re-identify a record. If an adversary does not have background knowledge of a variable then it cannot be a quasi-identifier. The manner in which an adversary can obtain such background knowledge will determine which attacks on a data set are plausible. For example, the background knowledge may be available because the adversary knows a particular target individual in the disclosed data set, an individual in the data set has a visible characteristic that is also described in the data set, or the background knowledge exists in a public or semi-public registry.

Examples of quasi-identifiers include sex, date of birth or age, language spoken at home, ethnic origin, marital status, criminal history, total income, visible minority status, profession, event dates (such as admission, discharge, procedure, death, specimen collection, visit/encounter), codes (such as diagnosis codes, procedure codes, and adverse event codes), country of birth, birth weight, and birth plurality.

Step 2: Mask (transform) direct identifiers

Once the direct identifiers have been determined, masking techniques must be applied to those direct identifiers. Relevant masking techniques include the following: (a) removal of the direct identifiers, and (b) replacement of the unique direct identifiers with pseudonyms. Once masking is completed there is virtually no risk of re-identification from direct identifiers.

Step 3: Perform threat modelling

Threat modelling consists of the identification of the plausible adversaries and what information they may be able to access.

Step 4: Determine minimal acceptable data utility

It is important to determine in advance the minimal relevant data based on the quasi-identifiers. This is essentially an examination of what fields are considered to be the most appropriate given the purpose of the use or provision of access to the researcher in this case. This step concludes with the imposition of practical limits on how some data may be de-identified and the analyses that may need to be performed later on. For a non-public data set a minimal requirement may be to allow the replication of another published study.

Step 5: Determine re-identification risk threshold

What constitutes acceptable risk? As an outcome of the process used to define the threshold, the security and privacy controls that need to be imposed on the Qualified Investigator, if any, are defined.

Step 6: Import (sample) data from the source database

Importing the data from the source database may be a simple or complex exercise, depending on the data model of the source data set. This step is included explicitly in the process because it can consume significant resources and must be accounted for in any planning of de-identification.

Step 7: Evaluate actual re-identification risk

The actual risk is computed from the data set using the appropriate measure of re-identification probability (e.g., maximum and average risk). A number of parameters need to be set to compute risk, such as the sampling fraction.

Step 8: Compare the actual risk with the threshold

Compare the actual risk with the threshold determined in Step 5.

Step 9: Set parameters and apply data transformations

If the measured risk is higher than the threshold then further de-identification and masking methods are applied to the data such as generalization, suppression, randomization, and sub-sampling.

Sometimes a solution cannot be found within the specified parameters, and it is necessary to go back and reset the parameters. It may also be necessary to modify the threshold and adjust some of the assumptions behind the original risk assessment. Alternatively, some of the assumptions about acceptable data utility may need to be renegotiated with the Qualified Investigator.

Step 10: Perform diagnostics on the solution

If the measured risk is lower than the threshold then diagnostics should be performed on the solution. Diagnostics may be objective or subjective.

An objective diagnostic would evaluate the sensitivity of the de-identification solution to violations of assumptions that were made. For example, an assumption may be that an adversary might know a secondary diagnosis of a participant, or if there is uncertainty about the sampling fraction of the data set then a sensitivity to that value can be performed. A subjective diagnostic would determine whether the utility of the data is sufficiently high for the intended purposes of the use or disclosure.

If the diagnostics are satisfactory, then the de-identified data are exported and a report documenting the de-identification, as well as the conditions accompanying the de-identification, is produced. On the other hand, if the diagnostics are not satisfactory, the re-identification parameters may need to be modified, the risk threshold may have to be adjusted, and the original assumptions about minimal, acceptable utility renegotiated with the data user.

If the data custodian is providing access to multiple de-identifications of the same data set to different Qualified Investigators, then there is a risk of an adversary trying to link these different data sets to get a more detailed data set. There are two ways for the data custodian to manage this risk. First, the data sharing agreement should prohibit linking data sets without permission. Second, the data custodian can pre-emptively compute the risk of re-identification if different versions of the data set are linked and apply further perturbations to the data accordingly.

Step 11: Export transformed data to external data set

Exporting the de-identified data to the destination database may be a simple or complex exercise, depending on the data model of the destination database. This step is included explicitly in the process because it can consume significant resources and must be accounted for in any planning of de-identification.

Step 12: Produce report documenting process and results

At the end of the de-identification process, a report documenting the process and results is produced and provided to the data custodian."

(El Emam, 2013b)

Assessments of the Council of Canadian Academies

The assessment reports listed below are accessible through the Council's website (www.scienceadvice.ca):

- Accessing Health and Health-Related Data in Canada (2015)
- Leading in the Digital World: Opportunities for Canada's Memory Institutions (2015)
- Policing Canada in the 21st Century: New Policing for New Challenges (2014)
- Energy Prices and Business Decision-Making in Canada: Preparing for the Energy Future (2014)
- Improving Medicines for Children in Canada (2014)
- Science Culture: Where Canada Stands (2014)
- Enabling Sustainability in an Interconnected World (2014)
- Environmental Impacts of Shale Gas Extraction in Canada (2014)
- Aboriginal Food Security in Northern Canada: An Assessment of the State of Knowledge (2014)
- Ocean Science in Canada: Meeting the Challenge, Seizing the Opportunity (2013)
- The Health Effects of Conducted Energy Weapons (2013)
- The State of Industrial R&D in Canada (2013)
- Innovation Impacts: Measurement and Assessment (2013)
- Water and Agriculture in Canada: Towards Sustainable Management of Water Resources (2013)
- Strengthening Canada's Research Capacity: The Gender Dimension (2012)
- The State of Science and Technology in Canada (2012)
- Informing Research Choices: Indicators and Judgment (2012)
- Integrating Emerging Technologies into Chemical Safety Assessment (2012)
- Healthy Animals, Healthy Canada (2011)
- Canadian Taxonomy: Exploring Biodiversity, Creating Opportunity (2010)
- Honesty, Accountability, and Trust: Fostering Research Integrity in Canada (2010)
- Better Research for Better Business (2009)
- The Sustainable Management of Groundwater in Canada (2009)
- Innovation and Business Strategy: Why Canada Falls Short (2009)
- Vision for the Canadian Arctic Research Initiative: Assessing the Opportunities (2008)
- Energy from Gas Hydrates: Assessing the Opportunities and Challenges for Canada (2008)
- Small Is Different: A Science Perspective on the Regulatory Challenges of the Nanoscale (2008)

- Influenza and the Role of Personal Protective Respiratory Equipment: An Assessment of the Evidence (2007)
- The State of Science and Technology in Canada (2006)

The assessments listed below are in the process of expert panel deliberation:

- Wind Turbine Noise and Human Health
- STEM Skills for the Future
- The Potential for New and Emerging Technologies to Reduce the Environmental Impacts of Oil Sands Development
- RISK: Is the Message Getting Through?
- Energy Use and Climate Change: A Synthesis of the Latest Evidence

Board of Governors of the Council of Canadian Academies*

Margaret Bloodworth, C.M., Chair, Former Federal Deputy Minister and National Security Advisor (Ottawa, ON)

Graham Bell, FRSC, President, Royal Society of Canada; Research Director, James McGill Professor, Chair, Department of Biology, McGill University (Montréal, QC)

John Cairns, FCAHS, President, Canadian Academy of Health Sciences; Professor of Medicine, University of British Columbia (Vancouver, BC)

Henry Friesen, C.C., FRSC, FCAHS, Vice Chair, Distinguished Professor Emeritus and Senior Fellow, Centre for the Advancement of Medicine, Faculty of Medicine, University of Manitoba (Winnipeg, MB)

Carol P. Herbert, FCAHS, Professor of Family Medicine, Western University (London, ON)

Claude Jean, Executive Vice President and General Manager, Foundry Operation, Semiconductor, Teledyne DALSA (Bromont, QC)

Peter MacKinnon, O.C., Former President and Vice-Chancellor, University of Saskatchewan (Saskatoon, SK)

Jeremy McNeil, FRSC, Helen Battle Professor of Chemical Ecology, Western University (London, ON)

Axel Meisen, C.M., FCAE, Former Chair of Foresight at Alberta Innovates – Technology Futures (AITF) (Edmonton, AB)

Lydia Miljan, Associate Professor of Political Science and Chair of the Arts and Science Program, University of Windsor (Windsor, ON)

Ted Morton, Executive Fellow at the School of Public Policy, Professor of Political Science, University of Calgary (Calgary, AB)

P. Kim Sturgess, FCAE, CEO and Founder, Alberta WaterSMART (Calgary, AB)

* Affiliations as of December 2014

Scientific Advisory Committee of the Council of Canadian Academies*

Susan A. McDaniel, FRSC, Chair, Director, Prentice Institute; Canada Research Chair (Tier 1) in Global Population and Life Course; Prentice Research Chair in Global Population and Economy; Professor of Sociology, University of Lethbridge (Lethbridge, AB)

Lorne Babiuk, O.C., FRSC, FCAHS, Vice President (Research), University of Alberta (Edmonton, AB)

Murray S. Campbell, Senior Manager, AI and Optimization, IBM T.J. Watson Research Center (Yorktown Heights, NY)

Clarissa Desjardins, CEO, Clementia Pharmaceuticals Inc. (Montréal, QC)

Jean Gray, C.M., FCAHS, Professor of Medicine (Emeritus), Dalhousie University (Halifax, NS)

John Hepburn, FRSC, Vice-President, Research and International, University of British Columbia (Vancouver, BC)

Gregory S. Kealey, FRSC, Professor, Department of History, University of New Brunswick (Fredericton, NB)

Daniel Krewski, Professor of Epidemiology and Community Medicine and Scientific Director of the McLaughlin Centre for Population Health Risk Assessment, University of Ottawa (Ottawa, ON)

Avrim Lazar, Former President and CEO, Forest Products Association of Canada (Ottawa, ON)

Norbert Morgenstern, C.M., FRSC, FCAE, University Professor (Emeritus), Civil Engineering, University of Alberta (Edmonton, AB)

Sarah P. Otto, FRSC, Professor and Director of the Biodiversity Research Centre, University of British Columbia (Vancouver, BC)

*Affiliations as of December 2014

Council of Canadian Academies
Conseil des académies canadiennes

Council of Canadian Academies
180 Elgin Street, Suite 1401
Ottawa, ON K2P 2K3
Tel: 613-567-5000
www.scienceadvice.ca